

Cognitive recovery in acute stroke: Measurement and facilitation of change

Hannah Tehan

Bachelor of Psychology (Honours)

Submitted as part fulfilment of the requirements for the degree of

Masters of Psychology (Clinical)/Doctor of Philosophy.

School of Psychology,

Australian Catholic University.

June, 2015.

### Statement of Authorship and Sources

This thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No parts of this thesis have been submitted towards the award in any other tertiary institution. No other person's work has been used without due acknowledgment in the main text of the thesis. All research procedures reported in the thesis received the approval of the relevant Ethics Committee.

Name: Hannah Tehan

Date: 04/06/2015

Signed: 

### Acknowledgements

Completion of this thesis would not have been possible without the encouragement and support of the stroke patients who participated in my study, my supervisors, colleagues, family and friends.

My primary supervisor Anne Tolan who has shown me infinite support, who has guided me through this project, who has kept me company in my highs and lows, has never pretended that this was going to be easy, or doubted that I could do it. You are so appreciated. My secondary, but no less wonderful supervisor, Kate Witteveen who could remember the pain, but also knew the rewards. Kate, on more than one occasion your calm sense of reason and gentle warmth kept this project, and multiple people from falling apart. Thank you for coming to my presentations, they were better for having you there.

Thank you to my panel members at ACU (and USQ) who provided important guidance and advice to make my thesis the best it could be.

Thank you to Donna Ward, Donna Pinsker, Mia Mariani and the whole clinical neuropsychology team at The Prince Charles Hospital who took me in as one of their own, and supported me to get the project off the ground and keep it running. A further thank you to Dr Lan and Dr Chan and the staff at the Acute Stroke and Rehabilitation Ward at TPCCH for welcoming me while I collected data.

And an important thank you to the patients at TPCCH who agreed to participate in my study, to my friends and family who were guinea pigs for me, and for stretching their networks so I could have a good normative sample. This project would of course not be possible without people's generosity with their time and cognitive effort.

And lastly, thank you to my dad who, beyond the immeasurable support he has given me for this project every day for the last five years, has been teaching me every minute for the last twenty-six. Above sandcastles, 80s music, and an affinity for vests, you have taught me what strong, independent and fiercely intelligent women look like. You recognised them, supervised them, befriended them, and to your credit and astonishing luck, married one. And you shared them with me, thank you. You have demonstrated strength, compassion, generosity, of course style, and what it is to be a thoroughly decent human being, so that I can recognise and find them for myself. And I have.

**Contents**

<b>Statement of Authorship and Sources .....</b>	<b>ii</b>
<b>List of Tables .....</b>	<b>xx</b>
<b>List of Figures .....</b>	<b>xxiv</b>
<b>Abstract.....</b>	<b>1</b>
<b>Chapter 1: Stroke and the Neural Underpinnings of Spontaneous Recovery .....</b>	<b>4</b>
<b>1.1 Thesis Overview .....</b>	<b>4</b>
<b>1.2 Stroke: Types, Symptomatology and Diagnosis.....</b>	<b>9</b>
<b>1.3 Natural Recovery – Repair Mechanisms .....</b>	<b>12</b>
<b>1.3.1 Animal Studies .....</b>	<b>12</b>
1.3.1.1 Damage.....	13
1.3.1.2 Recovery.....	14
<b>1.3.2 Studies using Human Participants .....</b>	<b>15</b>
<b>1.4 Therapy .....</b>	<b>17</b>
<b>1.4.1 Types of Intervention .....</b>	<b>18</b>
1.4.1.1 Molecules and cell-based therapies.....	18
1.4.1.2 Electromagnetic stimulation .....	18
1.4.1.3 Behavioural interventions.....	19
<b>1.4.2 The Therapeutic Process.....</b>	<b>19</b>
1.4.2.1 Restoration, recruitment and retraining .....	19
1.4.2.2 Timing of interventions.....	23
<b>1.5 Chapter Summary .....</b>	<b>24</b>
<b>Chapter 2: Cognitive Recovery and Training after Stroke.....</b>	<b>27</b>

<b>2.1 Chapter Overview.....</b>	<b>27</b>
<b>2.2 Scoping out Cognitive Impairment .....</b>	<b>28</b>
<b>2.2.1 Breadth and Complexity.....</b>	<b>28</b>
<b>2.2.2 Approaches to Cognition .....</b>	<b>28</b>
<b>2.2.3 Cognitive Psychology and Interventions .....</b>	<b>32</b>
<b>2.3 Assessing Cognitive Impairment Following Stroke.....</b>	<b>34</b>
<b>2.3.1 Screening Tools .....</b>	<b>34</b>
2.3.1.1 Mini-Mental State Exam. ....	35
2.3.1.2 Montreal Cognitive Assessment. ....	38
<b>2.3.2 Neuropsychological Tests.....</b>	<b>39</b>
<b>2.4 Cognitive Domains affected by Stroke .....</b>	<b>42</b>
<b>2.4.1 Acute Phase .....</b>	<b>43</b>
<b>2.4.2 Chronic Phase .....</b>	<b>48</b>
<b>2.5 Natural Recovery of Cognitive Functions.....</b>	<b>53</b>
<b>2.5.1 Recovery in the Chronic Phase.....</b>	<b>54</b>
<b>2.5.2 Recovery in the Acute Phase.....</b>	<b>60</b>
<b>2.5.3 Section Summary .....</b>	<b>63</b>
<b>2.6 Cognitive Training.....</b>	<b>65</b>
<b>2.6.1 Cognitive Training in the Chronic Phase following Stroke.....</b>	<b>68</b>
<b>2.6.2 Section Summary .....</b>	<b>74</b>
<b>2.6.3 Cognitive Training in the Acute Phase following Stroke.....</b>	<b>76</b>
<b>2.7 Recovery versus Training in Chronic Phase.....</b>	<b>76</b>
<b>2.8 Characteristics of Effective Training Programs.....</b>	<b>79</b>
<b>2.9 Literature Review Conclusions and Aims of Study.....</b>	<b>83</b>

<b>2.9.1 Conclusions</b> .....	<b>83</b>
<b>2.9.2 Aims and Objectives</b> . ....	<b>87</b>
<b>Chapter 3: Development of Neuropsychological and Cognitive Tasks</b> .....	<b>92</b>
<b>3.1 Chapter Overview</b> .....	<b>92</b>
<b>3.2 Neuropsychological Tests</b> .....	<b>94</b>
<b>3.3 Developing the Cognitive Tasks</b> .....	<b>97</b>
<b>3.3.1 Associative Networks</b> .....	<b>100</b>
<b>3.3.2 Working Memory and Dual Task Performance</b> .....	<b>105</b>
<b>3.4 Cognitive Tasks</b> .....	<b>109</b>
<b>3.5 Task Battery</b> .....	<b>113</b>
<b>3.5.1 Stroop Colour Word Task</b> .....	<b>113</b>
<b>3.5.2 Digit Span</b> .....	<b>114</b>
<b>3.5.3 Rey Tangled Lines Test</b> .....	<b>115</b>
<b>3.5.4 Phonemic and Semantic Verbal Fluency Tests</b> .....	<b>116</b>
3.5.4.1 Similarities. ....	117
3.5.4.2 Differences.....	118
3.5.4.3 The two processes approach .....	120
3.5.4.4. Phonemic fluency.....	123
3.5.4.5 Semantic fluency.....	123
<b>3.5.5 Common Associates Test</b> .....	<b>123</b>
3.5.5.1 Primary task.....	124
3.5.5.2 Dual task component.....	126
3.5.5.3 Scoring.....	126
<b>3.5.6 Immediate Serial Recall</b> .....	<b>127</b>

3.5.6.1 Primary task.....	127
<b>3.5.6.2 Dual task component.....</b>	<b>129</b>
3.5.6.3 Scoring.....	130
<b>3.5.7 The Snap Task.....</b>	<b>130</b>
3.5.7.1 Primary task.....	130
3.5.7.2 Dual task component.....	131
3.5.7.4 Scoring.....	132
<b>3.5.8 The Anagram Task.....</b>	<b>132</b>
3.5.8.1 Primary task.....	133
3.5.8.2 Dual task component.....	134
3.5.8.3 Scoring.....	135
<b>3.6 Testing Protocol.....</b>	<b>135</b>
<b>3.7 Chapter Summary.....</b>	<b>137</b>
<b>Chapter 4: Evaluation of the Cognitive Tasks.....</b>	<b>140</b>
<b>4.1 Chapter Overview.....</b>	<b>140</b>
<b>4.2 Experiment 4.1A.....</b>	<b>144</b>
<b>4.3 Method.....</b>	<b>145</b>
4.3.1 Ethics.....	145
4.3.2 Participants.....	145
4.3.3 Materials.....	145
4.3.4 Procedure.....	145
<b>4.4 Results.....</b>	<b>146</b>
4.4.1 Response Production.....	146
4.4.2 Reliability.....	147

<b>4.5 Discussion</b> .....	<b>149</b>
<b>4.4 Experiment 4.1B</b> .....	<b>149</b>
<b>4.5 Method</b> .....	<b>149</b>
<b>4.5.1 Participants</b> .....	<b>149</b>
<b>4.5.2 Materials</b> .....	<b>149</b>
<b>4.5.3 Procedure</b> .....	<b>149</b>
<b>4.6 Results and Discussion</b> .....	<b>150</b>
<b>4.6.1 Production of the Associate</b> .....	<b>150</b>
<b>4.6.2 Single Event Monitoring (Test 1)</b> .....	<b>152</b>
<b>4.6.3 Dual-Event Time-Switching (Test 2)</b> .....	<b>153</b>
<b>4.6.4 Reliability</b> .....	<b>154</b>
<b>4.7 General Discussion of Experiment 4.1</b> .....	<b>155</b>
<b>4.8 Experiment 4.2</b> .....	<b>156</b>
<b>4.9 Method</b> .....	<b>157</b>
<b>4.9.1 Participants</b> .....	<b>157</b>
<b>4.9.2 Materials</b> .....	<b>157</b>
<b>4.9.3 Procedure</b> .....	<b>157</b>
<b>4.10 Results</b> .....	<b>158</b>
<b>4.10.1 Snap Task</b> .....	<b>158</b>
4.10.1.1 Repetition and benchmarks.....	158
4.10.1.2 Reliability.....	158
<b>4.10.2. Anagram Task</b> .....	<b>160</b>
4.10.2.1 Repetition and benchmarks.....	160
4.10.2.2 Reliability.....	161

<b>4.10.3 Immediate Serial Recall Task</b> .....	<b>162</b>
4.10.3.1 Repetition and benchmarks.....	162
4.10.3.2 Reliability.....	163
<b>4.11 Discussion</b> .....	<b>165</b>
<b>4.12 General Discussion</b> .....	<b>166</b>
<b>4.13 Chapter Summary</b> .....	<b>169</b>
<b>Chapter 5: Normative Data for the Full Battery</b> .....	<b>170</b>
<b>5.1 Chapter Overview</b> .....	<b>170</b>
<b>5.2 Characteristics of Normative Group</b> .....	<b>171</b>
<b>5.3 Evaluating Improvement</b> .....	<b>173</b>
<b>5.3.1 Slopes and Intercepts</b> .....	<b>173</b>
<b>5.3.2 Temporal stability confidence intervals</b> .....	<b>176</b>
<b>5.4 Experiment 5.1</b> .....	<b>179</b>
<b>5.5 Method</b> .....	<b>180</b>
<b>5.5.1. Ethics</b> .....	<b>180</b>
<b>5.5.2 Participants</b> .....	<b>180</b>
<b>5.5.3 Materials and Procedure</b> .....	<b>180</b>
<b>5.6 Results</b> .....	<b>182</b>
<b>5.6.1 Snap Task</b> .....	<b>182</b>
5.6.1.1 Repetition.....	182
5.6.1.2 Pair detection and dual task detection. ....	183
5.6.1.3 Slopes and intercepts.....	183
5.6.1.4 Temporal stability confidence intervals. ....	183

5.6.1.5 Reliability. ....	184
<b>5.6.2 Anagram Task.....</b>	<b>185</b>
5.6.2.1 Repetition.....	185
5.6.2.2 Anagram errors and dual task detection. ....	186
5.6.2.3 Repetition and Benchmarks. ....	187
5.6.2.4 Slopes and intercepts. ....	188
5.6.2.5 Temporal stability confidence intervals. ....	188
5.6.2.6 Reliability. ....	188
<b>5.6.3 Stroop Colour Word Task.....</b>	<b>189</b>
5.6.3.1 Repetition.....	189
5.6.3.2 Slopes and intercepts.....	191
5.6.3.3 Temporal stability confidence intervals. ....	191
5.6.3.4 Reliability. ....	191
<b>5.6.4 Digit Span.....</b>	<b>193</b>
5.6.4.1 Repetition and benchmarks.....	193
5.6.4.2. Slopes and intercepts.....	194
5.6.4.3 Temporal stability confidence intervals. ....	194
5.6.4.4 Reliability. ....	194
<b>5.6.5 Rey Tangled Lines.....</b>	<b>196</b>
5.6.5.1 Repetition.....	196
5.6.5.2 Slopes and intercepts.....	197
5.6.5.3 Temporal stability confidence intervals. ....	197
5.6.5.4 Reliability. ....	197
<b>5.6.6 Phonemic Fluency Test.....</b>	<b>198</b>
5.6.6.1 Repetition.....	199

5.6.6.2 Slopes and intercepts.....	200
5.6.6.3 Temporal stability confidence intervals.....	200
5.6.6.4 Reliability.....	200
<b>5.6.7 Semantic Fluency Test .....</b>	<b>203</b>
5.6.7.1 Repetition.....	203
5.6.7.2 Slopes and intercepts.....	204
5.6.7.3 Temporal stability confidence intervals.....	204
5.6.7.4 Reliability.....	204
<b>5.7 Discussion.....</b>	<b>206</b>
5.7.1 Repetition Effects.....	207
5.7.2 Dual-Task and Benchmark Effects.....	209
5.7.3 Psychometric Properties.....	210
5.7.4 Demographic Effects.....	211
<b>5.8 Chapter Summary .....</b>	<b>214</b>
<b>Chapter 6: Introduction to Case Studies .....</b>	<b>217</b>
<b>6.1 Introductory Issues.....</b>	<b>217</b>
6.1.1 Heterogeneity.....	217
6.1.2 Comorbidities.....	217
6.1.3 Normative Sample versus Matched Controls.....	219
6.1.4 Evaluating Learning.....	220
6.1.4.1 Individual Tasks.....	220
6.1.4.2 Comparing initially impaired and unimpaired tasks.....	223
<b>6.2 Case Study Format .....</b>	<b>225</b>
6.2.1 Background.....	225

<b>6.2.2 Method</b> .....	226
<b>6.2.3 Results</b> .....	226
<b>6.2.4 Discussion and Overall Evaluation</b> .....	227
<b>Chapter 7 Case Studies: Neuropsychological Tasks</b> .....	228
<b>7.1 Introduction</b> .....	228
<b>7.2 Case Study 1: GL</b> .....	228
<b>7.2.1. Participant Background</b> .....	228
7.2.1.1 Precipitating event.....	228
7.2.1.2 Diagnosis. ....	228
7.2.1.3 Cognitive performance. ....	228
7.2.1.4 Other factors.....	229
<b>7.3 Method</b> .....	229
<b>7.3.1 Test Administered</b> .....	229
<b>7.3.2 Test Sessions</b> .....	229
<b>7.3.3 Test Notes</b> .....	230
<b>7.4 Results</b> .....	231
<b>7.4.1 Absolute Performance</b> .....	231
7.4.1.1 Initial levels of impairment.....	231
<b>7.4.2 Learning</b> .....	234
7.4.2.1 Temporal stability confidence intervals. ....	234
7.4.2.2 Slopes. ....	234
<b>7.4.3 Impaired and Unimpaired Tasks</b> .....	237
<b>7.5 Discussion</b> .....	239
<b>7.6 Case Study 2: MF</b> .....	240

<b>7.6.1 Participant Background.....</b>	<b>240</b>
7.6.1.1 Precipitating event.....	240
7.6.1.2 Diagnosis.....	240
7.6.1.3 Cognitive performance.....	240
7.6.1.4 Other factors.....	240
<b>7.7 Method .....</b>	<b>240</b>
<b>7.7.1 Test Administered .....</b>	<b>240</b>
<b>7.7.2 Test Sessions .....</b>	<b>241</b>
<b>7.7.3 Test Notes .....</b>	<b>242</b>
<b>7.8 Results .....</b>	<b>242</b>
<b>7.8.1 Absolute Performance.....</b>	<b>242</b>
7.8.1.1 Initial levels of impairment.....	242
<b>7.8.2 Learning .....</b>	<b>245</b>
7.8.2.1 Temporal stability confidence intervals.....	245
7.8.2.2 Slopes.....	245
<b>7.8.3 Impaired and Unimpaired Tasks .....</b>	<b>247</b>
<b>7.9 Discussion .....</b>	<b>249</b>
<b>7.10 Case Study 3: KS .....</b>	<b>251</b>
<b>7.10.1 Participant Background.....</b>	<b>251</b>
7.10.1.1 Precipitating event.....	251
7.10.1.2 Diagnosis.....	251
7.10.1.3 Cognitive performance.....	251
<b>7.11 Method .....</b>	<b>251</b>
<b>7.11.1 Test Administered .....</b>	<b>251</b>

<b>7.11.2 Test Sessions</b> .....	<b>252</b>
<b>7.11.3 Test Notes</b> .....	<b>253</b>
<b>7.12 Results</b> .....	<b>254</b>
<b>7.12.1 Absolute Performance</b> .....	<b>254</b>
7.12.1.1 Initial levels of impairment.....	254
<b>7.12.2 Learning</b> .....	<b>257</b>
7.12.2.1 Temporal stability confidence intervals. ....	257
7.12.2.2 Slopes.....	258
<b>7.12.3 Impaired and Unimpaired Tasks</b> .....	<b>260</b>
<b>7.13 Discussion</b> .....	<b>262</b>
<b>7.14 Case Study 4: LL</b> .....	<b>263</b>
<b>7.14.1 Participant Background</b> .....	<b>263</b>
7.14.1.2 Diagnosis. ....	263
<b>7.15 Method</b> .....	<b>263</b>
<b>7.15.1 Test Administered</b> .....	<b>263</b>
<b>7.15.2 Test Sessions</b> .....	<b>264</b>
<b>7.15.3 Test Notes</b> .....	<b>265</b>
<b>7.16 Results</b> .....	<b>265</b>
<b>7.16.1 Absolute Performance</b> .....	<b>265</b>
7.16.1.1 Initial levels of impairment.....	265
<b>7.16.2 Learning</b> .....	<b>266</b>
7.16.2.1 Temporal stability confidence intervals. ....	266
7.16.2.2 Slopes.....	270
<b>7.16.3 Impaired and Unimpaired Tasks</b> .....	<b>271</b>

<b>7.17 Discussion</b> .....	<b>272</b>
<b>7.18 Case Study 5: TB</b> .....	<b>273</b>
<b>7.18.1 Participant Background</b> .....	<b>273</b>
7.18.1.1 Precipitating event.....	273
7.18.1.2 Diagnosis. ....	273
7.18.1.3 Cognitive performance. ....	273
7.18.1.4 Other factors.....	273
<b>7.19 Method</b> .....	<b>273</b>
<b>7.19.1 Test Administered</b> .....	<b>273</b>
<b>7.19.2 Test Sessions</b> .....	<b>274</b>
<b>7.19.20 Test Notes</b> .....	<b>274</b>
<b>7.20 Results</b> .....	<b>276</b>
<b>7.20.1 Absolute Performance</b> .....	<b>276</b>
7.20.1.1 Initial impairments. ....	276
<b>7.20.2 Learning</b> .....	<b>278</b>
7.20.2.1 Temporal stability confidence intervals. ....	278
7.20.2.2 Slopes.....	279
<b>7.20.3 Impaired and Unimpaired Tests</b> .....	<b>281</b>
<b>7.21 Discussion</b> .....	<b>283</b>
<b>7.22 General Discussion</b> .....	<b>283</b>
<b>Chapter 8 Cognitive Tasks</b> .....	<b>287</b>
<b>8.1 Introduction</b> .....	<b>287</b>
<b>8.2 Case Study 1: GL</b> .....	<b>288</b>
<b>8.2.1. Participant Background</b> .....	<b>288</b>

<b>8.3 Method .....</b>	<b>288</b>
<b>8.3.1 Test Administered .....</b>	<b>288</b>
<b>8.4 Results .....</b>	<b>289</b>
<b>8.4.1 Initial Levels of Performance and Temporal Stability .....</b>	<b>289</b>
<b>8.4.2 Dual Task Performance. ....</b>	<b>290</b>
<b>8.4.3 Letter Length in the Anagram Task.....</b>	<b>293</b>
<b>8.5 Discussion and Overall Evaluation .....</b>	<b>293</b>
<b>8.6 Case Study 2: MF.....</b>	<b>294</b>
<b>8.6.1 Participant Background.....</b>	<b>294</b>
<b>8.7 Method .....</b>	<b>294</b>
<b>8.7.1 Test Administered .....</b>	<b>294</b>
<b>8.7.2 Test Notes .....</b>	<b>294</b>
<b>8.8 Results .....</b>	<b>295</b>
<b>8.8.1 Initial Levels of Performance and Temporal Stability .....</b>	<b>295</b>
<b>8.8.2 Dual Task Performance .....</b>	<b>296</b>
<b>8.8.3 Letter Length in the Anagram Task.....</b>	<b>297</b>
<b>8.9 Discussion and Overall Evaluation .....</b>	<b>297</b>
<b>8.10 Case Study 3: KS .....</b>	<b>297</b>
<b>8.10.1 Participant Background.....</b>	<b>297</b>
<b>8.11 Method .....</b>	<b>298</b>
<b>8.11.1 Test Administered .....</b>	<b>298</b>
<b>8.11.2 Test Notes .....</b>	<b>298</b>
<b>8.12 Results .....</b>	<b>298</b>

8.12.1 Initial Levels of Performance and Temporal Stability .....	298
8.12.2 Dual Task Performance .....	300
8.12.3 Letter Length in the Anagram Task.....	300
8.13 Discussion and Overall Evaluation .....	301
8.14 Case Study 4: LL .....	301
8.14.1 Participant Background.....	301
8.15 Method .....	302
8.15.1 Test Administered .....	302
8.15.2 Test Notes .....	302
8.16 Results .....	303
8.16.1 Initial Levels of Performance and Temporal Stability .....	303
8.16.2 Dual Task Performance .....	304
8.16.3 Letter Length in the Anagram Task.....	304
8.17 Discussion and Overall Evaluation .....	305
8.18 Comparing Patients .....	305
8.19 General Discussion.....	309
Chapter 9: Evaluation of Project .....	311
9.1 Chapter Overview.....	311
9.2 Rationale for the Study.....	311
9.3 Outcomes.....	313
9.3.1 Neuropsychological Tests.....	313
9.3.1.1 Initial levels of impairment.....	313
9.3.1.2 Tasks affected by stroke. ....	314

9.3.1.3 Learning.....	316
<b>9.3.2 Cognitive Tasks .....</b>	<b>318</b>
9.3.2.1 Learning.....	319
9.3.2.2 Dual task effects.....	320
9.3.2.3 Benchmark effects.....	321
<b>9.4 Theoretical Implications.....</b>	<b>321</b>
<b>9.4.1 Assessment.....</b>	<b>321</b>
<b>9.4.2 Learning .....</b>	<b>328</b>
9.4.2.1 Quality of data.....	328
9.4.2.2 Facilitating biological recovery. ....	330
9.4.2.3 Working memory training. ....	330
<b>9.5 Limitations .....</b>	<b>333</b>
<b>9.5.1 Issues Associated with Levels of Impairment.....</b>	<b>333</b>
<b>9.5.2 Issues Associated with Learning.....</b>	<b>336</b>
<b>9.5.2 Issues Associated with Sample Size .....</b>	<b>338</b>
<b>9.6 Future Studies.....</b>	<b>340</b>
<b>9.7 Conclusions .....</b>	<b>342</b>
<b>References .....</b>	<b>345</b>
<b>Appendix A .....</b>	<b>375</b>
<b>Human Research Ethics Approval Forms .....</b>	<b>375</b>
<b>Appendix B .....</b>	<b>379</b>
<b>Information Letter and Consent Instructions used in Experiments 4.1, 4.2. and</b>	
<b>5.1.....</b>	<b>379</b>

<b>Appendix C:.....</b>	<b>382</b>
<b>Information Letter and Consent Instructions used in Case Studies.....</b>	<b>382</b>

### List of Tables

Table 2.1	<i>Cognitive Domains and the Tests Associated with those Domains</i>	4 1
Table 3.1	<i>Neuropsychological Tests, their Administration Time and Underpinning Cognitive Domains used in the Test Battery.</i>	9 5
Table 3.2	<i>Cognitive Tasks, Administration Time and Underlying Cognitive Domains of Tests used the Battery</i>	1 09
Table 4.1	<i>Pearson's Correlation Coefficients for Target Production on Common Associates Test</i>	1 47
Table 4.2	<i>Intraclass Correlations [95% Confidence Intervals] on Target Production of the Common Associates Test</i>	1 47
Table 4.3	<i>Pearson's Correlation Coefficients for Target Production on Common Associates Test</i>	1 54
Table 4.4	<i>Intraclass Correlations [95% Confidence Intervals] for Target Production on the Common Associates Test</i>	1 54
Table 4.5	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Snap Task</i>	1 58
Table 4.6	<i>Intraclass Correlations [95% Confidence Intervals] on the Snap Task</i>	1 58
Table 4.7	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Anagram Task</i>	1 60
Table 4.8	<i>Intraclass Correlations [95% Confidence Intervals] on the Anagram Task</i>	1 61
Table 4.9	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Related Lists in Immediate Serial Recall.</i>	1 63
Table 4.10	<i>Intraclass Correlations [95% Confidence Intervals] on the Immediate Serial Recall Task</i>	1 64
Table 5.1	<i>Tasks Administered in the Test Battery across Testing Sessions</i>	1 80
Table 5.2	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Snap Task</i>	1 83
Table 5.3	<i>Intraclass Correlations [95% Confidence Intervals] for the Snap Task</i>	1 83
Table 5.4	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Anagram Task</i>	1 88
Table 5.5	<i>Intraclass Correlations [95% Confidence Intervals] for the Anagram Task</i>	1 88
Table 5.6	<i>Intraclass Correlations [95% Confidence Intervals] for the sub-tests of the Stroop task</i>	1 90
Table 5.7	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Read Sub-test of the Stroop task</i>	1 91
Table 5.8	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Colour Naming Sub-test of the Stroop task</i>	1 91
Table 5.9	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Colour-Word Sub-test of the Stroop task</i>	1 92
Table 5.10	<i>Intraclass Correlations [95% Confidence Intervals] for Forward and Backward Digit Span</i>	1 94
Table 5.11	<i>Pearson's Correlation Coefficients for Six Testing Sessions on Forward Span</i>	1 94

Table 5.12	<i>Pearson's Correlation Coefficients for Six Testing Sessions on Backward Span</i>	1 95
Table 5.13	<i>Intraclass Correlations [95% Confidence Intervals] for Rey Tangled Lines Test</i>	1 96
Table 5.14	<i>Internal Consistency and Pearson's Correlation Coefficients for Six Testing Sessions on the Rey Tangled Lines test</i>	1 97
Table 5.15	<i>Intraclass Correlations [95% Confidence Intervals] for the Three Scores on the Phonemic Fluency Task</i>	1 99
Table 5.16	<i>Pearson's Correlation Coefficients for First Responses on Phonemic Fluency Task</i>	2 00
Table 5.17	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Later Responses in the Phonemic Fluency Task</i>	2 00
Table 5.18	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Total Responses in the Phonemic Fluency Task</i>	2 01
Table 5.19	<i>Intraclass Correlations for the Three Scores on the Semantic Fluency Task</i>	2 03
Table 5.20	<i>Pearson's Correlation Coefficients for First Responses on Semantic Fluency Task</i>	2 03
Table 5.21	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Later Responses in the Semantic Fluency Task</i>	2 04
Table 5.22	<i>Pearson's Correlation Coefficients for Six Testing Sessions on the Total Responses in the Semantic Fluency Task</i>	2 04
Table 5.23	<i>Demographic Effects on Performance of Neuropsychological and Cognitive Tests</i>	2 10
Table 5.23	<i>Summary of Task Characteristics</i>	2 12
Table 7.1	<i>Tests Administered to GL across Six Test Occasions.</i>	2 27
Table 7.2	<i>Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions</i>	2 28
Table 7.3	<i>GL's Performance across Tasks expressed in z-scores</i>	2 30
Table 7.4	<i>GL's Learning Slopes relative to Normative Sample</i>	2 34
Table 7.5	<i>Summary of Performance Outcomes for GL</i>	2 35
Table 7.6	<i>Tests Administered to MF across Five Test Occasions.</i>	2 38
Table 7.7	<i>Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions</i>	2 38
Table 7.8	<i>MF's Performance across Tasks expressed in z-scores</i>	2 40
Table 7.9	<i>MF's Learning Slopes relative to Normative Sample</i>	2 44
Table 7.10	<i>Summary of Performance Outcomes for MF</i>	2 45
Table 7.11	<i>Tests Administered to KS across five Test Occasions</i>	2 49
Table 7.12	<i>Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions</i>	2 49

Table 7.13	<i>KS's Performance across Tasks expressed in z-scores</i>	2
		52
Table 7.14	<i>KS's Learning Slopes relative to Normative Sample</i>	2
		57
Table 7.15	<i>Summary of Performance Outcomes for KS</i>	2
		58
Table 7.16	<i>Tests Administered to LL across Six Test Occasions.</i>	2
		61
Table 7.17	<i>Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions</i>	2
		61
Table 7.18	<i>LL's Performance across Tasks expressed in z-scores</i>	2
		65
Table 7.19	<i>LL's Learning Slopes relative to Normative Sample</i>	2
		67
Table 7.20	<i>Summary of Performance Outcomes for LL</i>	2
		68
Table 7.21	<i>Tests Administered to TB across Five Test Occasions.</i>	2
		71
Table 7.22	<i>Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions</i>	2
		71
Table 7.23	<i>TB's Performance across Tasks expressed in z-scores</i>	2
		75
Table 7.24	<i>TB's Learning Slopes relative to Normative Sample</i>	2
		78
Table 7.25	<i>Summary of Performance Outcomes for TB</i>	2
		79
Table 8.1	<i>Cognitive Tests Administered to GL across Six Test Occasions.</i>	2
		86
Table 8.2	<i>GL's Performance on the Cognitive Tasks, Expressed in Z-scores, across Test Sessions</i>	2
		87
Table 8.3	<i>Cognitive Tests Administered to MF across Six Test Occasions.</i>	2
		90
Table 8.4	<i>MF's Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions</i>	2
		92
Table 8.5	<i>Cognitive Tests Administered to KS across Six Test Occasions.</i>	2
		94
Table 8.6	<i>KS's Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions</i>	2
		95
Table 8.7	<i>Cognitive Tests Administered to LL across Six Test Occasions.</i>	2
		98
Table 8.8	<i>LL's Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions</i>	2
		99
Table 9.1	<i>Ratios of Read/Colour-word Scores and Colour Naming/Colour-word Scores Averaged across Sessions</i>	3
		18
Table 9.2	<i>Ratios of Items Produced in the Later Response Period to First Response Period Averaged across Sessions</i>	3
		20
Table 9.3	<i>Ratios of Items Produced on the Semantic to Phonemic Fluency Tasks Averaged across Sessions</i>	3
		21



### List of Figures

<i>Figure 1.1.</i>	Temporal changes in cell functioning post-stroke. (Reproduced from Wieloch and Nikolich (2006) p. 259.)	1 3
<i>Figure 1.2.</i>	Changes in cortical motor map as a function of repetitive motor therapy. (Figure adapted from Kleim, 2011.)	2 3
<i>Figure 2.1.</i>	Frequency of test failure between control and stroke groups in Tatemichi et al. (1994).	4 4
<i>Figure 2.2.</i>	Prevalence of cognitive impairment at three stages in cognitive recovery. (Figure reproduced from Hurford et al., 2013).	6 2
<i>Figure 2.3.</i>	Training outcomes in Nordvik et al. (2012) study.	7 7
<i>Figure 2.4.</i>	Summary of prevalence, time-course of recovery, and initiation of training outcomes of the literature review.	8 4
<i>Figure 3.1.</i>	Stimulus for Rey Tangled Lines test.	1 14
<i>Figure 4.1.</i>	Target production on the three forms of the Common Associates Test. Error bars present the standard error of the mean.	1 46
<i>Figure 4.2.</i>	Performance on Associate Production on the Three Forms of the Common Associates Test. Error bars present the standard error of the mean.	1 50
<i>Figure 4.3.</i>	Detection of the capitalised word across the three forms of the Common Associates Test. Error bars present the standard error of the mean.	1 51
<i>Figure 4.4.</i>	Detection of events and time change across the three forms of the Common Associates Test. Error bars present the standard error of the mean.	1 53
<i>Figure 4.5.</i>	Completion time on the Snap task as a function of testing session. Error bars represent the standard error of the mean.	1 57
<i>Figure 4.6.</i>	Completion time on the Anagram task as a function of testing session and letter length. Error bars represent the standard error of the mean.	1 59
<i>Figure 4.7.</i>	Number of correctly recalled items on the Immediate Serial Recall task as a function of associative relatedness.	1 61
<i>Figure 5.1.</i>	Training outcomes (slopes) of four hypothetical participants.	1 74
<i>Figure 5.2.</i>	Training outcomes (slopes and intercepts) of four hypothetical participants.	1 75
<i>Figure 5.3.</i>	Learning effects based on temporal stability confidence intervals.	1 77
<i>Figure 5.4.</i>	Performance on the Snap task as a function of testing session. Error bars represent temporal stability 95% confidence intervals based on initial test performance.	1 81
<i>Figure 5.5.</i>	Pair detection (primary task) and card detection (secondary task) performance on the Snap task as a function of testing session. Error bars represent the standard error of the mean.	1 82
<i>Figure 5.6.</i>	Completion time on the Anagram task as a function of testing session. Error bars represent 95% confidence interval for repeated testing.	1 84

<i>Figure 5.7.</i>	Number of errors and number of dual task detections on the Anagram task as a function of testing session. Error bars represent standard error of the mean.	1 85
<i>Figure 5.8.</i>	Completion time on the Anagram task as a function of testing session and letter length. Error bars represent the standard error of the mean.	1 86
<i>Figure 5.9.</i>	Performance on the Stroop task as a function of test sessions.	1 89
<i>Figure 5.10.</i>	Performance on Digit Span task across test sessions.	1 92
<i>Figure 5.11.</i>	Performance on the Rey Tangled Lines test as a function of test session.	1 95
<i>Figure 5.12.</i>	Performance on the phonemic fluency task as a function of test sessions.	1 97
<i>Figure 5.13.</i>	Performance on the semantic fluency task as a function of test sessions.	2 01
<i>Figure 5.14.</i>	Learning effects expressed as scaled scores as function of test measure and test session.	2 13
<i>Figure 6.1.</i>	Performance of two hypothetical patients with stroke on a cognitive task as a function of test session.	2 19
<i>Figure 6.2.</i>	Slope values for control and patients with stroke on a hypothetical cognitive task.	2 20
<i>Figure 6.3.</i>	Overall summary of stroke patient's performance expressed in averaged scaled scores.	2 23
<i>Figure 7.1.</i>	GL's performance relative to normative sample.	2 31
<i>Figure 7.2.</i>	GL's slope values for all tasks compared to the normative sample.	2 33
<i>Figure 7.3.</i>	GL's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.	2 36
<i>Figure 7.4.</i>	MF's performance relative to normative sample.	2 41
<i>Figure 7.5.</i>	MF's slope values for all tasks compared to the normative sample.	2 43
<i>Figure 7.6.</i>	MF's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.	2 46
<i>Figure 7.7.</i>	KS's performance relative to normative sample.	2 53
<i>Figure 7.8.</i>	Slope values for KS relative to normative and control samples.	2 56
<i>Figure 7.9.</i>	KS's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.	2 59
<i>Figure 7.10.</i>	LL's performance relative to normative sample.	2 64
<i>Figure 7.11.</i>	LL's slope values for all tasks compared to the normative sample.	2 66
<i>Figure 7.12.</i>	LL's performance on initially impaired and initially	2

unimpaired tasks in comparison to normative and control groups.	69
<i>Figure 7.13.</i> TB's performance relative to normative sample and control group.	2 74
<i>Figure 7.14.</i> TB's slope values for all tasks compared to the normative sample.	2 77
<i>Figure 7.15.</i> TB's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.	2 80
<i>Figure 8.1.</i> GL's performance on snap and Anagram tasks as a function of testing session.	2 88
<i>Figure 8.2.</i> MF's performance the Anagram task as a function of testing session.	2 91
<i>Figure 8.3.</i> KS's performance the Anagram task as a function of testing session.	2 95
<i>Figure 8.4.</i> LL's performance the Anagram task as a function of testing session.	2 99
<i>Figure 8.5.</i> Patients with stroke' performance the Anagram task expressed in scaled scores as a function of dual task condition and testing session.	3 02

### **Abstract**

Strokes can affect any part of the brain and therefore have a wide range of potential outcomes including an array of cognitive deficits such as memory problems, neglect, problem solving difficulties and decision making errors. From a biological perspective, recovery from stroke can be categorized into two time phases, acute (up until 3 months) and chronic (3+ months), with most changes occurring in the acute phase. In the motor and speech areas, it is recognised that early intervention during the acute phase leads to the best long-term outcomes. The research into recovery of cognitive function is less well developed than in the motor and speech areas, however, there is a literature that explores the prevalence of cognitive impairments and recovery in the chronic phase. Such research is based upon patients with stroke' performance on batteries of standardised neuropsychological tests. This literature consistently demonstrates only small improvements over time. Training programs aimed at directly facilitating the recovery process, as opposed to developing compensatory behaviours to circumvent the effects of the impairment, have been implemented during the chronic phase. Many of these programs are based upon cognitive theories and employ commonly used cognitive psychology paradigms. These training programs have resulted in substantial improvements in the impaired functions. However, there are no studies that attempt to track changes in behaviour during the acute phase of stroke despite this being consistently demonstrated as a crucial period of recovery. The intent of the current research is to address this gap in the literature by exploring behaviour change in patients with stroke who are in hospital in the early stages of recovery from their first stroke.

The endpoint of the research was to compare the efficacy of neuropsychological and cognitive tests in assessing and/or promoting recovery. The neuropsychological tests in the battery were WAIS-III digit span, the Stroop Colour-word test, phonemic (COWAT) and semantic verbal fluency (Animal Naming plus to further trials) tasks, and the Rey Tangled

Lines Test. Four cognitive tasks were created specifically for the study, Common Associates Test, immediate serial recall, snap, and an anagram solving task, with all tasks employing a dual task methodology. Changes in behaviour were examined through six administrations of this battery over a two week period.

The empirical component of the thesis consisted of three phases. In the first phase the psychometric properties of the newly developed cognitive tests were examined using young healthy participants in order to establish adequate levels of reliability and to assess the degree to which practice effects were observed in the tests. The result of this phase was that two of the tests were rejected for failing to meet the specified criteria. The remaining two tests had good reliability and practice effects were present. In the second phase the battery was administered six times to a group of older participants (55-87years) across a two-week period. Interest lay in the extent to which practice effects were present in all tests and whether age, gender and education had any impact upon the strength of the practice effects. Furthermore, the intent was to confirm the psychometric properties of battery and to provide normative indicators of absolute performance and of learning (how behaviour improved with repetition). In the third phase the battery was administered to patients with stroke five or six times before they were discharged from hospital. Administration of the battery on the first test session permitted the identification of cognitive domains that had been impaired by the stroke and those that had not. Performance across the remaining sessions allowed for the examination of learning effects both at the individual task level, and at the level of tasks that were initially impaired and those that were initially not impaired. Five case studies are reported that showed that a) stable data could be obtained, b) the number of impaired tasks and the profile of those tasks differed among the patients with stroke, c) the degree of improvement varied among patients, c) improvement in unimpaired tasks was equivalent to that observed in the normative sample, d) practice effects could be distinguished from natural

recovery in the neuropsychological tests, and e) that cognitive tests produced more robust improvement than the simple repetition of the neuropsychological tests. The implication of the findings for the theoretical underpinnings of recovery from stroke and of clinical practice are discussed.

## **Chapter 1: Stroke and the Neural Underpinnings of Spontaneous Recovery**

### **1.1 Thesis Overview**

An early definition of stroke by Hatano (1976) is widely and encapsulates the important features of the diagnosis: “A stroke is a clinical syndrome characterised by rapidly developing clinical symptoms and/or signs of focal, and at times global loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin” (p. 541). Hatano’s definition gives a simple and clear description of a syndrome that is complex and heterogeneous in nature by way of areas of the brain that are affected and by diversity in symptomatology. The impact of stroke is evident given the statistic that stroke is the third most common cause of death in Australia (Lindley, 2008) and the most common cause of long-term disability in adults (Albert & Kesselring, 2012). The personal, societal and economic consequences of stroke are substantial.

Until relatively recently it was widely assumed that brain damage could not be repaired, although it was commonly recognised that those who survived an initial stroke showed some recovery of function in the first year post-stroke. In recent times, however, it has become apparent through both animal studies (Dancause et al., 2005; Johansson, 2004; Katsman, Zheng, Spinelli & Carmichael, 2003) and imaging studies involving human participants (Heiss et al., 1992; Marshall et al., 2000; Ward & Cohen, 2004) that there are many changes that occur in the brain after a traumatic injury. These changes are often grouped together under the banner of “neural plasticity”, but the basic premise is that changes to the brain are constantly happening and in the realm of stroke these changes underpin natural recovery. Furthermore, the timeframe for these changes appeared to operate on a negatively accelerating curve such that most of the biological changes and behavioural

recovery occur in the first few weeks and months after the stroke (Wieloch & Nikolich, 2006).

While many of the changes occur in a negatively accelerating curve across time, many researchers divide such a curve into discrete portions, sometimes two, acute and chronic, and sometimes three, acute, subacute, and chronic. Within each version, the events during the first eight to twelve weeks post-stroke appear to be crucial (Cassidy, Lewis, & Gray, 1998; Duncan et al. 1992; Kertesz, 1988). The current thesis adopts a two-stage model where acute and subacute phases are collapsed into one period. Consequently, the first twelve weeks are commonly referred to as the acute phase of recovery and the period after this is referred to as the chronic phase. Changes in the brain following trauma imply that interventions that assist in recovery are possible (Kleim, 2011) and that the most opportune time for an intervention is in the early days and weeks following the stroke (Cumming et al., 2011; Marin et al., 2003; Nudo, 2006).

The interest in plasticity changes has extended beyond the realm of stroke and other forms of brain damage. There has been an explosion in “brain training” programs aimed at developing cognitive abilities in normal children and adults alike. These programs have been developed largely to improve cognitive performance in maths, reading, and problem solving skills, as well as in general increase in cognitive processing. Numerous commercial programs now exist (e.g., CogMed, Cognifit, Lumosity, Jungle Memory) that claim to produce: beneficial effects in populations known to have cognitive issues, educational benefits for those who do not have any obvious impairment, increases in cognitive capacity and in intelligence, and the ability to offset normal age-related cognitive declines (cfr Melby-Lervåg & Hulme, 2012; Shipstead, Redick & Engle, 2012). While many of these commercial products have not been extensively evaluated (CogMed [[www.cogmed.com](http://www.cogmed.com)] is an exception),

it is becoming clear that repeated testing on many of the tasks in these products can improve performance on those tasks.

The current thesis is developed against this background. In Australia, when a person suffers a stroke and is admitted to hospital and subsequently commences rehabilitation in a hospital ward, it is frequently the case that they participate in a range of interventions. Physiotherapy is started if there are mobility issues, or speech therapy is commenced if there are language or swallowing issues, and occupational therapists are involved if it is anticipated that there will be functional living issues on discharge. In these areas, the effects of stroke are readily apparent. Because many cognitive difficulties are not as obvious (Planton et al., 2012), it is rarely the case that psychologists are involved in cognitive rehabilitation in this acute phase of recovery.

This lack of emphasis on cognitive impairment in the acute phase is reflected in the literature in two ways. Much of the extant literature has been concerned with identifying the extent of cognitive impairments, and it is clear using neuropsychological assessment procedures that such impairments are frequent outcomes of stroke (Tatemichi et al, 1994). Such deficits are more prevalent in the acute phase than the chronic phase (Hurford et al., 2013) and imply that cognitive recovery is possible. However, there is very little literature that documents the recovery process itself. That is, there are no published studies that explicitly examine changes in cognitive performance across the first weeks after a stroke. Consequently, the first aim of the thesis is to explore the assessment of cognitive impairment in the early days after stroke and to track the changes in executive functions, working memory, speed of processing, visual processing, and verbal fluency, during this critical period.

The second aspect of the lack of emphasis on cognitive rehabilitation is more direct. The literature involving cognitive rehabilitation and related training programs indicates that

most cognitive rehabilitation programs are only initiated months after the stroke had occurred. No published literature could be found where cognitive interventions were initiated in the hospital context in the days and weeks following a stroke. This represents not only a large gap in the literature, but also represents an area where clinical psychological practice could produce long-term benefits. Consequently, the second aim of the thesis is concerned with the process involved in creating cognitive interventions that could be commenced in the acute phase of recovery. The cognitive training literature indicates that cognitive skills do improve with training. However, interpreting exactly what improves with training is difficult to establish unequivocally, and the case is made that disentangling training effects from natural recovery effects in patients with stroke is even more difficult. As a precursor to developing an intervention program, the research explores the changes in performance when a series of neuropsychological tests are repeatedly administered and when a series of cognitive tasks are administered. Neuropsychological tasks have been designed with cognitive assessment in mind rather than serving as tools for an intervention. On the other hand, cognitive tasks used in intervention studies have generally been adopted from mainstream cognitive psychology and have been selected specifically to enhance performance. The current research aims to determine whether one class of tests is more sensitive to, or more potent at producing improvements in performance among patients with stroke and would thus serve as better candidates in the development of future interventions that could be implemented in the acute phase of stroke.

A case-study approach was deemed to be the most appropriate way to explore both the assessment of recovery and the comparison of cognitive and neuropsychological tests. To this end a battery tasks (neuropsychology and cognitive psychology) was developed and the battery was repeatedly administered over a two to three week period to five patients who had

recently suffered a stroke, and who had been admitted to the rehabilitation ward of a metropolitan hospital in Australia.

In Chapter 1 the effects of stroke, the process of natural recovery, and the range of therapies that have been developed to facilitate brain plasticity are described. This literature is based primarily in the context of motor and speech problems associated with stroke because it is these areas that have involved the greatest degree of research. The aim of this chapter is to explore the biological and behavioural changes that occur in the weeks and months post-stroke and to examine how both biological and behavioural effects are influenced by therapeutic interventions.

Chapter 2 discusses the issue of the cognitive impairments following a stroke: what cognitive factors are affected by stroke, how any such deficits are assessed and what such assessment practices suggest in terms of rehabilitation processes. The outcomes of this chapter suggest that a) multiple cognitive impairments are common outcomes of stroke, b) natural recovery occurs with most changes occurring in the early phase of recovery, and c) there is no literature available concerning interventions in the early phases post-stroke. The chapter also explores cognitive training issues that establish the need to explore differential changes in behaviour on neuropsychological and cognitive tasks, as a first step in creating effective interventions.

Chapter 3 introduces the neuropsychological and cognitive tests in the battery and provides the rationale behind task selection. The tasks used to track recovery were standard neuropsychological tests and these tasks were always administered in an identical format. Several tasks were developed to form the intervention component of the research. These tasks were derived from mainstream cognitive psychology and were based on notions of implicit priming through spreading activation among associative networks, and on cognitive load

considerations that are derived from dual task methodologies. The working memory training literature influenced a number of aspects of the training regime.

Because the cognitive tasks do not have a track history concerning learning effects or psychometric properties, Chapter 4 evaluated the utility of the cognitive tasks. The tasks were administered to a group of young, healthy participants to establish the feasibility of using these tasks and for determining the psychometric properties of the various tasks. To preview the results, two of the planned tasks failed to meet criteria for acceptance and were deleted from the subsequent battery.

Chapter 5 describes the outcomes of a study where the battery was administered a number of times to a group of healthy older participants. These data served as normative data by which levels of cognitive impairment could be derived for the patients with stroke. The data also provided normative information concerning changes in performance (learning) across testing sessions.

Chapter 6 provides a brief overview of the structure of case studies. It also outlines the various methods used to evaluate cognitive impairment and learning outcomes in both sets of tasks. Chapter 7 examines performance of the five patients with stroke on the neuropsychological tasks with the data being presented in 5 case studies. Chapter 8 does the same with the cognitive tasks.

In Chapter 9, the final chapter, the outcomes are reviewed and the tasks in the battery are evaluated for their continued utility. The implications and limitations of the research are discussed, and ideas for follow up research are provided.

## **1.2 Stroke: Types, Symptomatology and Diagnosis**

Hatano's (1976) definition of stroke is basically a cause and effect account. The cause is highly specific in that it involves a vascular insult on the brain and the effect is a loss of cerebral functioning that manifests itself in observable behaviour. More simply, a stroke is a

“brain attack”, caused by a disturbance in the brain’s blood supply leading to loss of cerebral function.

The least problematic form of a stroke is known as a transient ischaemic attack (TIA). Like other forms of stroke a TIA results from an inadequate blood supply to the brain, often caused by a blood clot or a haematological (blood) disease (Lindley, 2008). The blood disturbance in a TIA resolves itself within 24 hours, therefore making it a “transient attack”. A TIA can cause acute loss of cerebral functions but this also often resolves itself in a short period of time. TIAs are not considered any further in this thesis.

More serious forms of stroke are the ischaemic stroke and the haemorrhagic strokes. These forms can lead to permanent brain damage including paralysis and cognitive impairment (Lindley, 2008). An ischaemic stroke is a blockage/clot in the brain or in a blood vessel leading to the brain, which inhibits the blood supply to the brain resulting in the death or decreased functioning in the cells in the affected area. Four types of blood disturbance are thought to contribute to ischaemic stroke: a) thrombosis, a blood clot forming locally in the arteries that supply blood to the brain, b) embolism, a clot forming elsewhere in the body (usually heart or lungs) and travelling to the brain, c) systemic hypoperfusion, a general decrease in blood supply to the whole body (and therefore the brain) that is caused by poor cardiac functioning, for example heart attack, and d) venous thrombosis, a clot in the dural venous sinus that drains blood from the brain.

A haemorrhagic stroke is an accumulation of blood within the skull cavity. It is caused by damage to a blood vessel in the brain, potentially bleeding on to the tissue and also starving the brain of blood elsewhere (Lindley, 2008). A haemorrhagic stroke is categorised as either an intra-axial haemorrhage, which is an accumulation of blood inside the brain, or an extra-axial haemorrhage, which is blood inside the skull, but outside the brain, for example in the epidural/subdural space.

Ischaemic strokes, being caused by clots, are often treated with medicines that break down the clot (thrombolytic therapy) or by thinning the blood. Excess bleeding through haemorrhagic strokes is often treated with blood clotting agents. The opposing treatment methods highlight the importance of an accurate diagnosis.

Strokes can affect any part of the brain and therefore have a wide range of potential outcomes. These can include motor impairments such as paralysis, balance problems and muscle weakness, loss or partial loss of visual abilities, sensory impairment, speech and swallowing difficulties, as well as an array of cognitive deficits including memory problems, problem solving difficulties and decision making errors. The signs and symptoms of stroke vary depending on which part of the brain is affected. However, it is generally true that signs and symptoms appear suddenly, last between seconds and minutes and can have common elements. The heterogeneous nature of stroke means that the clinical symptoms of stroke vary significantly between patients. Furthermore, the symptoms often mimic other illnesses (Lindley, 2008). For example, numbness of limbs, weakness of limbs, dysarthria, aphasia, vertigo, loss of vision and ataxia, can individually be signs and symptoms of various other illnesses including epilepsy and various cancers (Lindley, 2008).

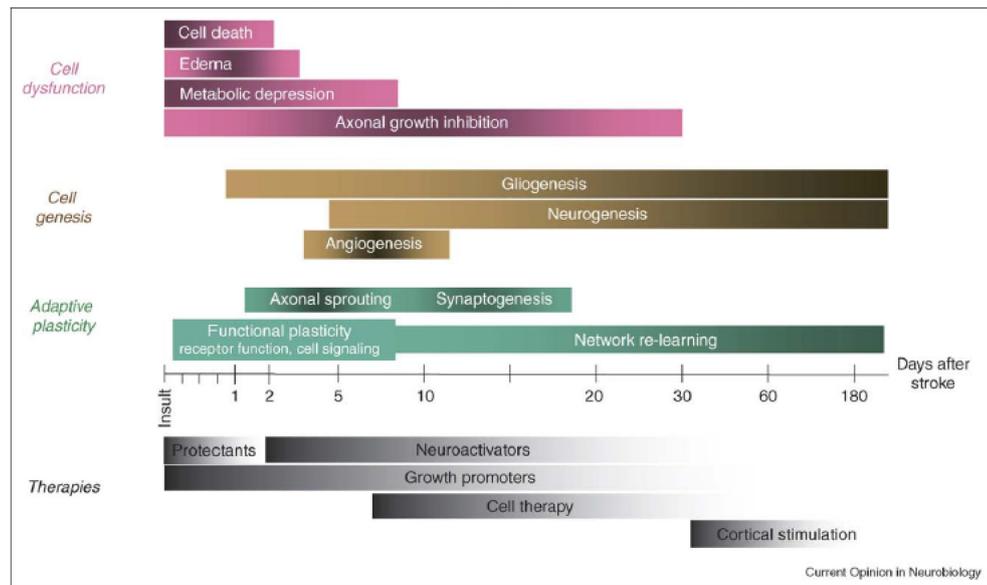
Problems in diagnosis are exacerbated by the fact that there is a limited time frame for some therapeutic outcomes. Patients who are diagnosed and treated with thinning or clotting medication within a 3-hour window have a significantly greater chance of recovery than those who are not treated within this time period (Lees et al., 2010). Research has indicated that patients who have had an ischaemic stroke and have had access to thrombolytic therapy within three hours of symptom presentation, show a lasting 12% improvement in recovery outcomes in comparison to those who did not receive therapy (The National Institute of Neurological Disorders and Stroke, 1995). After the 3-hour window, treatment outcomes are much less substantial (Hacke et al., 2008).

### **1.3 Natural Recovery – Repair Mechanisms**

Following a stroke, there is a period where damaged functions show some recovery, either partial or whole. Our current understanding of the recovery process and the biological changes that underpin it is derived from animal and human studies. In the case of animal studies, lesions to the vascular system in the brain are frequently induced in the motor cortex because neural damage can be controlled, and the subsequent changes in motor behaviour are relatively easy to study. In the case of research with human participants the bulk of the stroke research has also involved problems in motor area (although there is much literature on the language effects of stroke, and lesser literature on cognitive effects). Thus, motor functioning is the backdrop to most of the research into understanding recovery and the interventions aimed at facilitating recovery. These processes are described in the current chapter and similar processes in cognitive recovery are addressed in the next chapter.

#### **1.3.1 Animal Studies**

Animal studies have been instrumental in providing information about the natural recovery process, partly because the infarcts can be localised to specific areas (typically involving the motor system). The animal studies have allowed controlled exploration of the molecular, cellular, neuronal and vascular changes that occur in the days following stroke. Such changes occur in parallel not only in the regions near the infarct (peri-infarct regions) but also in remote, but connected, regions. Figure 1.1 provides a visual overview of the damage and repair function.



An overview of the activated parallel processes and therapeutic approaches following brain injury, specifically stroke. The temporal sequence of events is shown along a semi-logarithmic schematic timeline of 180 days after injury. Darker shading highlights the maximum intensity of the specific mechanism. Processes that are detrimental towards recovery are shown in pink. Processes of cell genesis are shown in brown, whereas those that underlie adaptive plasticity are shown in green. Prospective therapies that focus on neuroprotection and enhancement of regeneration and functional recovery are shown in gray.

*Figure 1.1.* Temporal changes in cell functioning post-stroke. (Reproduced from Wieloch and Nikolich (2006) p. 259.)

**1.3.1.1 Damage.** The death of neurons is one obvious source of damage brought about by vascular insult. However, damage is not limited to the death of affected neurons. Surviving tissue in the peri-infarct region can be affected by the presence of blood in the environs (oedema), and by pressure brought about by inflammation. The result of the stroke is that there is a reduction in metabolic processes. The processes that support growth in axons are inhibited, and dendritic and axonal damage can occur (Wielock & Nikolich, 2006). While the catabolic damage is usually repaired quickly (Katsman, Zheng, Spinelli & Carmichael, 2003), resultant structural damage can still produce abnormal functioning.

Damage is not confined to the areas surrounding the insult. It is apparent that other undamaged areas of the brain that have pathways connected to the damaged site also undergo substantial change as reflected in changes in blood flow or metabolism (Binkofski et al.,

1996; Feeney & Baron, 1986). The reduced functioning at undamaged sites is known as diaschisis.

**1.3.1.2 Recovery.** As Figure 1.1 indicates, there are repair processes that commence soon after the stroke has occurred. These processes can be considered as support processes and plasticity processes. In terms of the support processes, there is a marked increase in the generation of glia cells that are responsible for supporting neurones and their environment. The commencement of repair in the vascular system (angiogenesis) starts within three days of the stroke.

Direct changes to neural functioning occur through blocking of growth inhibitors, and these changes thereby facilitate structural changes in the neuron and changes in the links between neurons. Therefore, the size and structure of dendrite trees, and the type and number of receptor binding sites can change dramatically, and changes in the form of the axon spine and axonal growth are also apparent (Johansson, 2004; Schallert, Leasure & Kolb, 2000). New synapses can be formed where they previously did not exist. Axonal sprouting between areas that are not normally connected can lead to new pathways being created where none existed before the event (Dancause et al, 2005).

Figure 1.1 also provides an overview of the timeframe over which these events occur. During the first 10 days post-stroke most of the processes are initiated. Some processes peak early and then diminish and others persist for much longer periods. Wieloch and Nikolich (2006) argue for three distinct phases during the recovery process. The early phase of recovery is characterised by the reversal of diaschisis and activation of cell repair. The second phase, between weeks two and four involves the changes that occur in existing neural pathways. The third phase, which can occur over the time frame of years, involves the processes associated with creating new neuronal connections. Successful outcomes at each

stage are predictors of long-term recovery, suggesting that early intervention might be crucial in determining behavioural recovery from the effects of stroke.

### **1.3.2 Studies using Human Participants**

Studies of the damage caused by stroke and the subsequent recovery process in humans requires less invasive techniques than those used in animal studies. In addition, in human studies a wider range of pathologies can be studied in addition to changes in motor behaviour (though it is still the case that a majority of human studies are concerned with changes in the motor system). Electroencephalography (EEG) and imaging techniques are currently used to explore brain changes, with the latter becoming more and more sophisticated. These imaging techniques, for the most part, measure brain activity rather than the molecular and cellular changes that are the remit of the animal studies. The imaging techniques have also been applied in animal studies and the results show that there are many similar outcomes between humans and animals (Chouinard, Leonard, & Paus, 2006; van der Zijden et al., 2007).

As is the case in the animal studies, changes in brain activity associated with recovery have been noted in local, peri-infarct regions as well as in remote regions. In terms of local effects, it is clear that immediately post-stroke there is marked reduction in cortical activity, but such activity increases over time (Marshall et al., 2000; Nhan et al., 2003), with the degree of increase being a good predictor of behavioural recovery (Ward & Cohen, 2004). While increases in neural activities in peri-infarct areas have also been noted (Heiss et al., 1992), it has not yet been firmly established that the degree of activity in these regions is related to behavioural outcomes (Cramer et al., 2006).

In terms of effects on remote areas, imaging evidence tends to show three ways in which areas that are remote from, but richly connected to the stroke site, respond to the insult. The first involves activity in remote parts of the network, the second involves increased

activity in the contralateral hemisphere and the third involves changes in somatotopic maps in the cortex.

As noted earlier, one of the effects of stroke is the phenomenon of diaschisis, the reduced functioning of remote areas that are connected to the stroke site. In the cases where diaschisis has not occurred or has been resolved, there is frequently increased activation in the unaffected cortical areas of the network (e.g., motor, language, attention and visual pathways) in which the infarcted area is involved (Brion, Demeurisse, & Capon, 1989).

A decrease in neural activity in the affected area is often accompanied by increased activity in the same areas on the undamaged contralateral hemisphere, a pattern that is seen in other forms of brain insult besides stroke. Cross-hemispheric changes in activity result in reduced laterality in activity (Seitz et al., 1998). The precise reason for this shift in activity is unknown and may simply be another example of increased activity at remote sites or it may involve reduced inhibition from the affected site (Seitz et al., 1998). Others, however, have suggested that this increased activity may well be a sign that the contralateral hemisphere is taking on the function of the damaged areas (but see Palmer, Ashby & Hajek, 1992, for disconfirming data).

While the precise functions of both forms of activity at remote sites might not be fully understood these functions are reliably linked to behavioural outcomes. These changes in activity levels are time-dependent in that they peak and then decline afterwards. The rate of decline is predictive of behavioural outcomes with those showing the greatest declines also showing the best behavioural outcomes (Ward, Brown, Thompson, & Frackowiak, 2003).

Imaging studies show that following stroke there are changes in the degree of grey matter thickness in both peri-infarct areas and in contralateral areas. Basically, the notion is that other areas of the brain take on the function of the damaged areas by establishing new connections and new growth in these unaffected areas and these changes can result in cortical

thickness. For example, damage in the primary motor cortex induces increased recruitment of secondary pre-motor and supplementary motor cortex (Ward & Cohen, 2004), with the result that somatotopic maps change during recovery to reflect the reorganisation of neural functions (Nudo & Milliken, 1996).

### **1.3.3 Timeframe for Behavioural Recovery**

The timeframe associated with restorative biological processes is characterised by rapid molecular and cellular changes followed by neuroplasticity changes in terms of synaptogenesis, dendrite branching, and neuronal sprouting that result in changes in cortical thickness and topography. These changes are reflected in behavioural outcomes that occur within much the same time frame. With motor damage following stroke the most dramatic changes can be observed within the first 30 days post-stroke (Duncan et al., 1992, 1994; Wade et al, 1983). With language deficits the bulk of whatever improvement occurs is observed within the first three months (Kertesz, 1988; Kertesz & McCabe, 1977), as is the case with visual neglect (Cassidy, Lewis, & Gray, 1998; Hier, Mondlock, & Caplan, 1983). However, while behavioural recovery is most pronounced in the first month or so improvements can still be observed in the months (and sometimes years) following strokes but such gains are more modest (Cramer, 2008).

The behavioural and biological studies converge on a common time frame where a transition point appears between one and three months post-stroke. Behavioural changes are most marked in the acute phase of stroke with fewer and more modest gains during the chronic phase of stroke.

## **1.4 Therapy**

The previous section in this chapter describes the processes that occur in recovery where there is little in the way of intervention. In the following sections a) the types of therapeutic interventions; and b) the therapeutic processes will be described. The basic

assumption being made is that such interventions build on and augment the processes that underpin natural recovery. Again, much of the supporting literature is based on damage to the motor system.

#### **1.4.1 Types of Intervention**

The first medical interventions involve attempts to stem the causal problem via the administration of clotting agents or anti-coagulant medications. Subsequent interventions reflect the recovery process in many ways in that interventions are targeted at specific parts of the recovery process.

**1.4.1.1 Molecules and cell-based therapies.** The early stages of recovery are marked by a number of repair events and many substances have been examined as potential facilitators of the recovery process. Many of these interventions have been aimed at altering dysfunctional neurotransmitter systems (Berthier et al., 2006). Other substances that encourage growth in both vascular repair or in the cells that support repair, or in the changes in peri-infarct functioning, have also been successful in some instances (Cramer, 2008). Newer techniques involving cell-based therapies are also being trialled; for example the administration of neural stem cells. This line of research is still in its infancy (Cramer, 2008; 2011).

**1.4.1.2 Electromagnetic stimulation.** The recovery process is characterised by differences in cortical activity at specific sites and by cross hemispheric changes in the balance of activation. Transcranial magnetic stimulation and transcranial direct current stimulation are two non-invasive applications that can modulate brain activity. In both types, stimulation of specific areas of the cortex can be implemented such that targeted interventions are possible. Stimulation can either target areas that are underactive (Khedr, Ahmed, Fathy & Rothwell, 2005) or suppress activation in areas that are over active (Fregni et al, 2006) such that hemispheric balance is restored. Functional improvements of 10% in a

single session and 20% to 30% improvements over multiple sessions have been reported (Harris-Love & Cohen, 2006; Huang, Edwards, Rounis, Bhatia & Rothwell, 2005).

**1.4.1.3 Behavioural interventions.** A number of behavioural interventions have been employed to alter the levels of brain activity in one or other of the hemispheres. For motor problems, constraint induced movement therapy is being used increasingly frequently (Albert & Kesselring, 2012). The idea behind constraint induced therapy is that by constraining the unaffected limb the affected limb is forced into producing activity in the affected pathways. Such programs have shown clinically meaningful improvements that have lasted two years post-stroke (Wolf et al., 2006; Wolf et al., 2008). Similar techniques have successfully been applied to aphasia (Meinzer et al., 2007).

By far the most widely used behavioural intervention involves simple repetition of a task. Given that most human learning involves repeated exposure to and experience with the learning object, it is not surprising that such repetition features in almost all behavioural therapies. The effects of repetition per se are often overlooked. In terms of the constraint induced therapy, it is not the constraint that is producing the beneficial outcomes, but it is the repeated use of the affected limb that is producing the changes in activation and the improvement in functional outcome (Albert & Kesselring, 2012). The importance of repetitive experience cannot be overstated in any form of learning or training whether it is a difficult motor task like driving a golf ball in a straight line or an equally difficult cognitive task like learning statistics. Automatic tasks like walking are based on hours and hours of repetitive practice where failure was just as likely an outcome as success in the early phases of mastering the skill.

## **1.4.2 The Therapeutic Process**

**1.4.2.1 Restoration, recruitment and retraining.** Events in the brain during recovery appear to be very similar to those involved in the normal acquisition of new skills

(Kleim, 2011). Historically, much of the biology-based knowledge of plasticity and regeneration has come from studies investigating healthy brains of animals during repetitive learning tasks. Kleim (2011) has extensively studied brain density in rats that were trained on both simple and complex motor tasks. He reported that in comparison to rats that consistently moved on a bare, flat runway (simple motor task), rats who trained on complex obstacle courses, including ladders and ropes (complex motor task), possessed more synapses within the motor-cortex (Kleim, Lussnig, Schwarz, Comery & Greenough, 1996), and cerebellum (Kleim, Pipitone, Czerlanis, & Greenough, 1998), as well as dendritic growth (Withers & Greenough, 1989), and synaptogenesis (Kleim, Barbay, Cooper, Hogg, Reidel & Remple, 2002). Kleim's research indicated that repeated practice in an enriched environment of challenging motor tasks encouraged reorganisation of synapses and growth in associated neural circuits.

Similar neural changes have been observed in human studies where years of practice are involved or even when short periods of practice are involved. For example, in the case of repeated practice over years, chess grand masters show differences in connectivity among brain regions compared to novice chess players (Duan et al., 2014) as do masters in other board games (Jung et al., 2013). Grey matter volumes in the areas of cortex associated with spatial memory are larger in taxi drivers than in the normal population (Maguire, Woollett & Spiers, 2006; Spiers & Maguire, 2006). In expert musicians, the primary hand motor area is more extended and less symmetrical (Amunts et al., 1998) and parts of the corpus callosum are also larger ( Schlaug, Jancke, Huang, Staiger & Steinmetz, 1995).

The time period for skill acquisition need not be in terms of years. With a little as three months of training, jugglers show differences in those areas of the cortex that are associated with processing of visual motion (Draganski et al, 2004). Training on simple motor tasks shows similar effects. A number of studies have investigated neural growth after

skilled training on a hand task (Pascual-Leone, Dang, Cohen, Brasil-Neto, Cammarota & Hallett, 1995), an ankle task (Perez, Lungholt, Nyborg & Nielsen, 2004) and a tongue task (Svensson, Romaniello, Arendt-Nielsen & Sessle, 2003) and all resulted in increased density in associated areas in the motor cortex. Kleim (2011) proposed that learning a skill, whether it is a motor skill or a cognitive skill leads to observable changes in the cortex. He presents this as a fundamental framework for neurorehabilitation in humans arguing that similar observable changes in the cortex can be observed in the aftermath of stroke.

After injury or damage to the brain, functional improvement can be framed as a re-learning process (Kleim, 2011). Traditional therapies after injury include physiotherapy, where movement is practiced and either fully or partially restored; speech therapy, where word formation, sentence construction and other skills are practiced and improved; and occupational therapy, where daily living tasks such as showering, sitting safely and moving safely in an environment are rehearsed. The guided practice of therapy aims to re-acquire a person's ability to perform activities, which were lost as a result of injury (Kleim, 2011). Kleim frames re-learning as a three-stage biological process that results in functional and behavioural improvement. The first stage, *restoration*, is based on the resolution of diaschisis, that is, the restoration of function in uninjured areas of the brain closely associated to the injury site by close proximity or through a neural network. Therapy aims to promote *restoration* of dysfunctional areas, by encouraging recovery that happens naturally, effectively restoring the cortex to a more functional state.

The second stage in the neurorehabilitation process is *recruitment*, which refers to enlisting non-damaged parts of the brain to both compensate for and activate damaged areas. For example, Crosson et al, (2005) examined neural recruitment in post-stroke aphasia patients who had damage to Broca's area (left hemisphere). They used a complex left-hand movement to activate the motor cortex in the right hemisphere, and therefore promote

lateralisation of language to the right hemisphere. Results of the study showed increased lateralisation in the frontal cortex and a significant improvement in language functioning.

The final process is *retraining* which is closely linked with *restoration* and *recruitment*, in that rehabilitation uses non-damaged parts of the brain to perform functions of damaged areas. In comparison to *recruitment*, however, *retraining* requires the brain to make a permanent shift in the area of function, and begin to learn and adapt to novel tasks. An example of this is commonly seen in rehabilitation clinics in the form of “restraint therapy”, where a functional limb is restrained, forcing the use of the opposite dysfunctional limb. This in turn promotes neural functions to make a permanent shift to other areas of the brain. Kleim (2011) reiterates the importance of plasticity based rehabilitation as it underpins the functional reorganisation the brain experiences during rehabilitation, and therefore actual behavioural and functional change for the patient.

Figure 1.2 provides an example of how the three processes result in structural change. In this instance, the effects of all three forms of the rehabilitation process have occurred in the peri-infarct region of the affected cortex. The figure shows the motor map of forelimb movement representations within a rat’s motor cortex. The upper panel shows the area in the motor cortex that is devoted to elbow and shoulder representations and wrist and digit representations. An Ischaemic stroke has subsequently been induced and the bottom panels reflect the state of the motor representations after two weeks post-stroke. In the case of the control rat, one assumes that there has been some natural recovery occurring. In the case of the rat that has received two weeks of motor rehabilitation, the therapy has produced more restoration of the damaged areas compared to the control rat (i.e., more green areas). One can also see that areas of the elbow map have also been recruited and retrained to support the wrist movement (i.e., blue areas that are now green). Note the bar graph shows that the changes are reflected by the number of synapses per neuron.

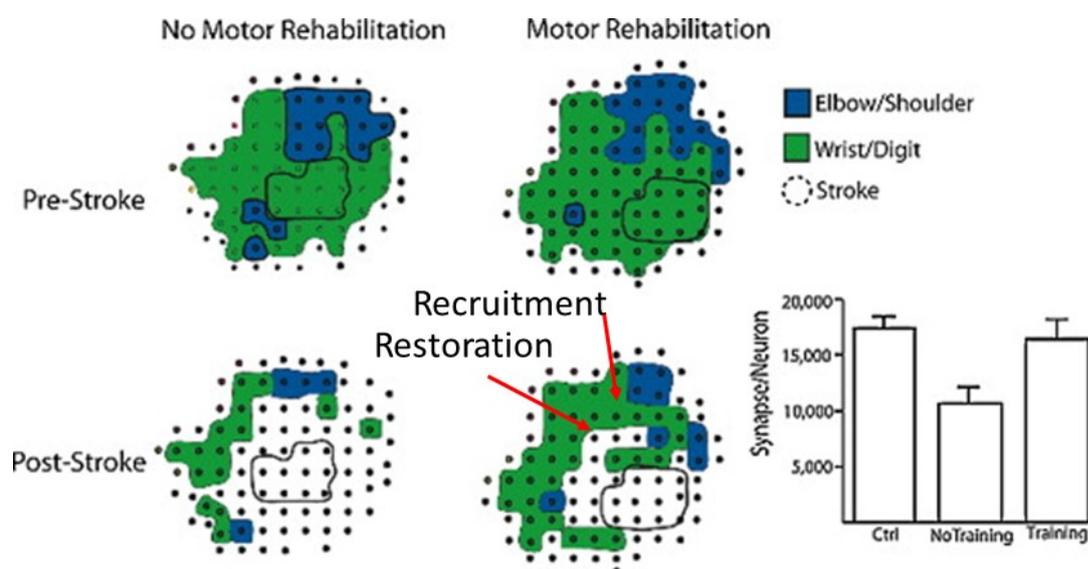


Figure 1.2. Changes in cortical motor map as a function of repetitive motor therapy. (Figure adapted from Kleim, 2011.)

**1.4.2.2 Timing of interventions.** As stated plasticity related events are initiated in the days following a stroke; they peak over the following three or four weeks and then decline over the following months. Behavioural recovery appears to follow a similar time frame with most gains being made in the early weeks after a stroke. It is not surprising that researchers have examined the timing of interventions to determine when such intervention has maximal effects.

Animal studies have directly assessed changes in plasticity and functional outcomes as a function of when interventions are initiated. Studies in which forced activity was initiated within 24 hours of a stroke actually exacerbated the detrimental effects of the stroke (Humm, Kozlowski, James, Gott, & Schallert, 1998). However, if training started after 5 days then both functional outcomes and dendritic sprouting increased relative to conditions where training commenced at day 14. Furthermore, both factors were more pronounced when training began at day 14 than when training began at day 30 (Biernaskie, Chernenko, & Corbett, 2004). Similar studies show that initiating active training during the first three weeks

after stroke leads to better behavioural and neuronal outcomes than inactivity during these periods (Marin, et al., 2003; Nudo, 2006; Wall & Egger, 1971).

Studies involving human patients with stroke with motor impairments show a similar pattern. Cumming et al. (2011) showed that mobilization within 24 hours of a stroke led to faster and better walking ability than groups for whom mobilization started later than this. Moreover, the advantage for the early mobilization group was maintained during the course of the study. Intervention in the early days following stroke has also resulted in shorter stays in hospital (Indredavik, Bakke, Slordahl, Rokseth, & Haheim, 1999), better functional outcomes (Maulden, Gassaway, Horn, Smout, & DeJong, 2005), reduction in disability and better quality of life outcomes (Musicco, Emberti, Nappi, & Caltagirone, 2003).

The results of both the animal and human studies converge on the importance of early intervention. It would appear that early mobilization of the damaged motor system leads to better behavioural outcomes and enhanced plasticity changes over and above those associated with natural recovery.

### **1.5 Chapter Summary**

This chapter examines the neurological outcomes of stroke and the subsequent behavioural changes that stem from the neural processes. There are different types of stroke based on causal factors, but all strokes can produce motor, language, perceptual and cognitive deficits. The degree and extent of deficits due to stroke can differ from person to person. Some strokes result in focal effects in that the damage is limited to specific neural mechanisms that have specific behavioural symptoms. Thus, people can experience a stroke that leads to paralysis or weakness in specific areas of the body; specific aspects of language can be affected; and neglect can be limited to specific parts of the visual field. Alternatively, quite diffuse areas can be affected in other types of stroke, such that there is widespread

damage across multiple systems. Chapter 2 will address this issue in more detail in terms of the cognitive effects of stroke.

When damage is caused by a stroke, repair mechanisms involving changes in molecular, cell, and activation functions come into play very soon afterwards. These mechanisms result in changes in the areas surrounding the infarct, remote areas that are networked to the infarct area, and changes in the unaffected hemisphere (Wieloch & Nikolich, 2006). These changes in the structures of the neurons, the links between neurons and the resultant changes in pathways are termed *neural plasticity*. Natural recovery is the global term used to convey the transformations that are characteristic of neural plasticity.

The review also indicates that the timeframe of natural recovery can be divided into two components: the acute phase where the repair mechanisms are initiated and peak in activity. Secondly, in the chronic phase, the repair mechanisms have peaked, but may continue to operate for much longer periods. The implications of this timeframe are that the bulk of recovery occurs during the acute phase (Cassidy, Lewis, & Gray, 1998; Duncan et al. 1992; Kertesz, 1988) and interventions are likely to have maximal effects if initiated during this period (Albert & Kesselring, 2012).

Therapeutic interventions have been devised to build on the recovery process. Pharmaceutical and other cell-based therapies are aimed at facilitating changes at the cellular level (Cramer, 2008). Stimulating or depressing cortical activity by external means has been used to both measure the degree of intact cortex, and to alter the level of cortical activity (Fregni et al., 2006; Khedr et al., 2005). Constraint induced therapy has been successfully used to reactivate damaged parts and pathways (Wolf et al., 2006; 2008). In fact, the therapeutic process can be conceived in terms of the restoration of affected areas, recruitment of other areas of the brain, and retraining of all areas (Kleim, 2011). The research indicates that with both animals and humans, repetitive practice underpins all facets of structural

changes (Albert & Kesselring, 2012). That is, to be able to restore, recruit and retrain, extensive repetitive practice is required. Therapy involves the same plasticity mechanisms that are involved in learning new tasks and skills. The distinction between acute and chronic phases of stroke is important because the beneficial outcomes of an intervention are maximized if the intervention is initiated within the acute phase of stroke (Cumming et al., 2011).

This chapter has relied heavily on animal and human studies that deal with recovery and therapy in the motor system. References have been made at times to similar findings involving focal cognitive deficits like aphasia and neglect. However, there are many other forms of cognitive processes that may be impaired by stroke. Chapter 2 explores more diffuse cognitive effects, and does so using the acute/chronic framework that appears to reflect key changes in biological restorative mechanisms and in motor recovery. If this framework holds for cognitive changes as well as for motor (and language, and visual neglect), then substantial and rapid changes should be observed in the early weeks post-stroke with less substantial changes occurring in the chronic phase of recovery. It should also be the case that interventions beginning in the acute phase should be more effective than those commencing in the chronic phase.

## **Chapter 2: Cognitive Recovery and Training after Stroke**

### **2.1 Chapter Overview**

In the previous chapter the neurological and behavioral processes involved in recovery after stroke were described in the context of damage to the motor system (for the most part). The recovery process was characterized by neural plasticity changes involving restoration of damaged areas and associated networks, recruitment of undamaged areas and retraining of recruited areas (Weiloch & Nikolich, 2006). Such changes meant that temporarily compromised systems regained function. The recovery process was most prominent during the first weeks after injury (acute phase) and subsequently stabilized during the following weeks, months and years (chronic phase) (Cassidy, Lewis, & Gray, 1998; Duncan et al. 1992; Kertesz, 1988). Therapeutic interventions aimed at building and facilitating the natural recovery process were also examined in Chapter 1. The evidence reviewed indicated that interventions were most effective, and the long-term beneficial effects of therapy were maximized, if the intervention was initiated during the acute phase of recovery (Albert & Kesselring, 2012). The distinction between acute and chronic phases of stroke thus provides a useful framework for exploring the brain's natural recovery process and the effectiveness of therapeutic interventions.

In Chapter 2 this framework is used to explore recovery and therapeutic interventions associated with cognitive impairments caused by stroke. Given the breadth of cognition and the different approaches taken towards cognition, this chapter begins by establishing limits to domains that will be the focus of the thesis. The ways in which cognitive performance is measured, and which domains are more or less sensitive to stroke are then considered. Finally, the literature concerning cognitive recovery and therapeutic interventions is reviewed.

## 2.2 Scoping out Cognitive Impairment

The study of cognitive impairments following stroke and the therapeutic response to such impairments are complicated by two broad factors: the breadth and complexity of cognition and the variety of approaches to cognition. Due to the magnitude of these factors, which are beyond the scope of a single thesis, it is necessary to limit the number of systems/domains examined and thereby provide a specific focus to the research.

### 2.2.1 Breadth and Complexity

In spite of the fact that stroke can have a deleterious impact on any cognitive task (Lezak, Howieson, Bigler & Tranel, 2012), not all cognitive impairments have received equal attention in the research literature. Cognitive disorders such as aphasia and visual neglect have been the subject of much research partly because these disorders are readily apparent to patients and have such a dramatic impact on behaviour. Having been the focus of a substantial body of research these disorders are *relatively* circumscribed and understood (Hurford, Charidimou, Fox, Cipolotti, & Werring, 2013). In contrast more subtle effects that come under the banner of cognitive impairment rather than specific disorders have received relatively little research. In 2007, Nys, van Zandvoort, de Kort, Jansen, de Haan and Kappelle observed "...not much is known about cognitive impairment and its determinants in the early phases of stroke" (p.409). As recently as 2013, Hurford et al. made the same point, "Furthermore, although major cortical functions (e.g., dysphasia, neglect) are routinely assessed in acute stroke, subtle cognitive impairments in the acute to subacute phase are largely ignored." (p. 238). These assertions identify a gap in the literature that the current thesis hopes to address.

### 2.2.2 Approaches to Cognition

Two broad approaches to cognition are relevant for current purposes. The neuropsychological approach emphasises the relationship between the structures of the brain

and behavior, and consequently focuses on how damage to particular structures of the brain results in changes in behavior (Kolb & Wishaw, 2009). At the clinical level this approach is represented in the way that cognitive abilities are assessed (Lezak et al., 2012). Cumming, Marshall and Lazar (2013) perhaps represent the current view of cognitive impairments following stroke when they observe that:

Cognition is not a unitary concept; it incorporates multiple domains, including attention (focusing, shifting, dividing or sustaining attention on a particular stimulus or task), executive function (planning, organising thoughts, inhibition, control), visuospatial ability (visual search, drawing, construction), memory (recall and recognition of visual and verbal information), and language (expressive and receptive). Classification is far from straightforward as domains are not independent. (p. 38)

In reducing cognition to manageable proportions, the domains and subdomains identified in the quotation are the ones that are adopted in the current thesis as they currently dominate the research literature regarding assessment of cognitive impairments following stroke. In short, the thesis deals primarily with the areas of executive function, attention, and selective aspects of memory, language and visuo-spatial processing.

Whereas the neuropsychological approach makes inferences concerning brain behavior relationships (Kolb & Wishaw, 2009; Lezak et al., 2012), the cognitive approach has historically not been constrained by neuroanatomical structures but rather accepts that cognitive processes are in fact mediated by the structures of the brain. Consequently, brain-behaviour models have no part in determining conclusions or predictions. The mainstream cognitive psychology approach is more interested in the structures, systems, and processes that determine normal cognitive performance rather than impaired functions (Ashcraft, 2005). From the perspective of cognitive psychology the Cumming et al. (2013) quote could

be reframed to include perception (top-down, bottom up); attention (selective, divided); visual imagery; memory (short-term and working memory, episodic memory, semantic memory, procedural memory, encoding and retrieval); knowledge (categories, semantic networks); language (word perception, sentence comprehension, text and story comprehension); problem solving; reasoning and decision making.

While cognitive psychology attempts to understand normal cognition, this approach does contribute to the stroke literature. Studies have explored the effects of stroke on working memory (Hoffman, Jefferies, Ehsan, Jones & Lambon Ralph, 2012; Vallat, Azouvi, Hardisson Meffert, Tessier, & Pradat-Diehl, 2005; van Geldorp, Kessels & Hendriks, 2013), on episodic memory (Dimitrov, Granetz, Peterson, Hollnagel, Alexander, & Grafman, 1999), on associative processing (Hales & Brewer, 2012), and categorization (Crutch & Warrington, 2008) to name but a few domains. Moreover, there are studies that involve the interaction among these systems, for example, working memory, associative memory, and language (Hoffman et al., 2012), or working memory and associative processing (van Geldorp et al., 2013). Both neuropsychological and cognitive psychology literatures need to be considered to understand the full range of the cognitive impairments caused by stroke.

To the extent that both approaches make many of the same assumptions they can be considered complementary. Both approaches make the same basic points that cognition is multi-faceted, that domains can be identified, that each domain has sub-domains and the domains are not independent of each other. However, in spite of the apparent similarities between the two approaches, there are differences between them that are problematic. The differences are most obvious in the sub-domains, where there is almost no overlap. For example in the area of memory, modality of presentation and retention interval feature heavily in the neuropsychological tests. In the Wechsler Memory Scales (WMS-IV), separate auditory and visual indices are derived for immediate memory and delayed memory

Moreover, there are specific tests that have been created to assess these modality-based domains. For example, the Benton Visual Retention test and Rey Complex Figure Test are two tests that evaluate visual memory. The California Verbal Learning Test and Rey Auditory Verbal Learning Test are two among many measures that assess auditory verbal memory. All four tests have an immediate memory test and a delayed memory test incorporated into the testing protocol to examine retention effects. In cognitive psychology, modality and retention effects on memory are important but sub-domains are not based on these factors, instead they are based on memory systems. Distinctions between procedural, semantic and episodic memory (Tulving, 1985), or long-term memory, short-term memory, and working memory (Baddeley, 1986, 2000) are more likely to be involved in classification of sub-domains.

Even in the instance where a system is common to both approaches, it is doubtful whether each approach is referring to the same thing. As an example, the Wechsler Memory Scales - III had a working memory index that was measured using letter-number sequencing, and forward and backward spatial span tasks with forward and backward digit span being included as optional tests. Working memory from this neuropsychological approach appears to involve the temporary storage of material for subsequent manipulation and processing. In the case of forward digit span and spatial span the processing is relatively simple (maintaining the order of items), whereas in backward digit and spatial spans and letter-number sequencing, items have to be maintained and then re-ordered.

In the cognition literature there are two distinct approaches to working memory. One approach, as typified by the research of Engle and his colleagues (Engle, Tuholski, Laughlin, & Conway, 1999; Unsworth & Engle, 2007a), views working memory along the same lines as the neuropsychology approach in that working memory is seen as a limited capacity system responsible for active maintenance and online manipulation of information over short

intervals (Unsworth & Engle, 2007a). However this approach makes a strong distinction between working memory and short-term memory systems, with the latter system involving passive storage without any processing component. They view forward and backward digit span and spatial span tasks as tests of short-term memory rather than tests of working memory and it is unclear whether they would see letter-number sequencing as another measure of short-term memory or a measure of working memory. The second and most popular account of working memory, Baddeley's working memory model (Baddeley & Hitch, 1974; Baddeley, 1986, 2000), depicts working memory not as a unitary construct but as a multi-component system involving an attentional allocation system (central executive) and a number of limited capacity storage buffers (the phonological loop, visuo-spatial sketch pad and episodic buffer). In this model, digit span would be handled by the phonological loop component, spatial span by the visuo-spatial sketch pad, and letter-number sequencing could be supported either by the phonological loop or by the episodic buffer. Even when the same construct is used in neuropsychological and cognitive approaches, there are very different meanings associated with the construct, and very different explanations for performance on the different tasks. A consequence of such imprecision is that a unified picture of the cognitive effects of stroke is unlikely to be formed.

One method of dealing with such confusion is to simply adopt one of the approaches and ignore the other. If the intent is simply to explore the natural recovery process of stroke then one could legitimately adopt the neuropsychological approach because its primary emphasis is on the assessment of cognitive ability in order to detect the presence or absence of cognitive impairments. However, if one is interested in interventions, cognitive psychology literature needs to be addressed because much of the intervention literature is based on intact cognitive systems and processes.

### **2.2.3 Cognitive Psychology and Interventions**

There is a long history of memory rehabilitation programs being informed by and based on theoretical perspectives on the different types of memory systems and memory processes (Wilson, 2009). Principles of errorless learning, distributed practice, enhancing retrieval cues, encoding specificity, levels of processing, and interactive visual imagery are just some of the theory-based factors that have been incorporated into rehabilitation programs (Grandmaison & Simard, 2003; Wilson, 2009). Simple theory-based interventions can often have immediate and dramatic effects. For example, it has long been established that forming an association between two items in paired-associates tasks helps a subject to remember those items (Humphreys, 1976). Furthermore, it has also been established that forming visual associations is a more effective strategy for remembering than forming verbal associations (Bower & Winzenz, 1970). Building on these robust memory effects, Hales and Brewer (2012) developed a rehabilitation task that explored the use of verbal or visual associative mnemonics for an episodic memory task. They describe the case of a person who developed aphasia after a left-hemisphere stroke, and were able to show that different brain regions were active for verbal and for visual forms of the task. The patient exhibited memory difficulties using her own existing (verbal) strategies and performance declined when she was asked to use a verbal associative strategy. The authors deemed this to be a natural consequence of the damage to her left hemisphere. However, when she was asked to use a visual associative strategy that they had shown activated areas in the right hemisphere, her performance on the memory task was error free. Thus, the principles and processes of memory (in this instance associations based on visual imagery), which were established in the 1970's, can still provide the basis of a simple and effective cognitive intervention today.

The role of cognitive psychology in interventions is wider than memory rehabilitation or determining the reasons for specific cognitive impairments. Recent developments in neural plasticity have been the driving force for the proliferation of cognitive training programs

based on working memory in particular and other executive tasks to a lesser extent (e.g., CogMed, Cognifit, Lumosity, Jungle Memory). In evaluating these programs cognitive psychologists working in the working memory domain have been instrumental in showing the methodological characteristics of a good cognitive training program, and the appropriate way to evaluate training effectiveness (Melby-Lervåg & Hulme, 2012; Shipstead, Redick & Engle, 2012). Most importantly, this literature reinforces the idea that good training programs need to be grounded in cognitive theory (Shipstead et al., 2012).

In sum, the effects of stroke on cognition can be explored from either the neuropsychological approach or the cognitive psychology approach. These approaches have different aims and assumptions, and even though both approaches are concerned with similar cognitive domains, the organisation of those domains and the tasks used to measure them can differ markedly. The neuropsychological approach emphasises brain-behaviour relationships and is interested in the cognitive system when it is impaired through some form of trauma. These aims reflect an emphasis on detecting and identifying cognitive impairments. The cognitive approach focuses more on the systems and processes that underpin normal cognition and cognitive models are typically independent of any brain behavior relationships. While these approaches have traditionally developed independently of each other, over recent decades interactions between both traditions have led to the emergence of the discipline of cognitive neuroscience. For current purposes, the cognitive approach is important because of its relevance to training interventions in the cognitive domains of interest, namely executive function, attention, and selective aspects of memory, language, and visuo-spatial processing.

## **2.3 Assessing Cognitive Impairment Following Stroke**

### **2.3.1 Screening Tools**

It is common practice in developed countries for cognitive assessments following stroke to occur in two phases (Planton et al., 2012). Firstly, cognitive screening is done in the

hospital setting in the acute phase of stroke. The screening tools are often brief short pen and paper tests that cover multiple cognitive domains using a single test for each domain. The screening tests typically take 10 minutes to administer but do not provide detailed indicators of impairment because the intent is simply to determine if an overall cognitive impairment exists. A low score on a screening tool indicates the probability of general impairment, but provides limited information about the type and severity of the deficit. That is, the screening tests identify that a cognitive problem exists, they do not necessarily identify where that problem lies.

Identifying the problem area occurs in the second phase through the administration of neuropsychological test batteries. A full neuropsychological test battery provides wide coverage of cognitive domains using multiple tests for each domain, but it is recommended (Lezak et al., 2012) that these tests not be administered until the patient has recovered sufficiently to provide stable, reliable data. Consequently, full neuropsychological batteries are typically administered in the chronic phase of recovery. Furthermore, they are often only administered when cognitive impairments have emerged and are causing functional disability or distress. They are not administered as a matter of routine.

The two most commonly used screening tests in the hospital setting are the Mini-Mental State Exam (MMSE) (Folstein, Folstein & McHugh, 1975) and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). Both screening tools offer a time and resource friendly assessment that is quick, easy to administer and results in an estimation of a patient's cognitive function (Nys, van Zandvoort, de Kort, Jansen, Kappelle, & de Haan, 2005a).

**2.3.1.1 Mini-Mental State Exam.** The MMSE is currently the most widely used cognitive screening tool, both in clinical practice and in medical research (Bour, Rasquin, Boreas, Limburg & Verhey, 2010; Nys et al., 2005a). It consists of 17 clinician-led questions

measuring the cognitive domains of orientation, registration, attention and calculation, recall, language and visuospatial functioning. The orientation domain assesses the patient's knowledge of the current date and their location; registration requires the patient to repeat three words chosen by the examiner; attention and calculation is measured by subtracting 7 from 100 repeatedly (serial 7s), or alternatively, to spell the word "*world*" backwards; and recall is measured by the patient's ability to remember the three words from the previous registration subtest. The language domain is assessed in more depth with five questions: To name items physically presented by the clinician; to repeat the sentence "*No ifs, ands, or buts*" (designed as a series of short words of low probability occurring together); to follow a simple three stage verbal command; to follow a visual command; and lastly, to write a grammatically correct sentence. Visuo-spatial ability is measured by the patient's ability to draw a copy of two intersecting pentagons. The test takes approximately 10 minutes to complete. It is scored out of a total of 30 and a score of 24 is the most widely used cut-off point to identify impairment. Brayne, Nickson, McCracken, Gill and Johnson, (1998) reported that scores on these tasks, with a particular emphasis on the attention and calculation subtests, are affected by education achievement, cultural environment and sensory impairment. The wide use of the MMSE for global screening has been justified on the basis of its psychometric properties, with research demonstrating acceptable levels of internal consistency, high levels of inter-rater reliability (Dick, Guilloff, Stewart, Blackstock, Bielawska, Paul, & Marsden, 1984), and its close relationship to traditional comprehensive tests of cognition, such as the Wechsler Adult Intelligence Scales within the normal population (Dick et al., 1984).

The MMSE was originally developed to screen for dementia and delirium in a psychiatric population, and as a measure of this, showing strong validity, good reliability estimates and high levels of sensitivity and specificity (Folstein, Folstein, & McHugh, 1975).

Since its development the MMSE has been applied in a wide variety of clinical populations. In spite of its wide use, there is debate concerning its utility particularly in regards to its sensitivity to specific types of cognitive deficits and specificity to a wide range of disorders (e.g., Appelros, 2005; O'Sullivan, Morris, & Markus, 2005; Tombaugh & McIntyre, 1992). Numerous studies have demonstrated poor predictive validity outside of dementia, especially in neurological populations and in psychiatric populations (Faustman, Moses, & Csernansky, 1990; Grace, Nadler, White, Guilmette, Giuliano, Monsch, et al., 1995). The sensitivity of the MMSE to identify cognitive impairment varies significantly, with estimates anywhere between .62 and .96 depending on the cut off score (Blake, McKinney, Treece, Lee, & Lincoln, 2002). For example, a cut-off score of 23/24 has successfully screened for dementia with a sensitivity of .96 and specificity of .83 (Bour et al., 2010). However, a cut-off score of 24, when screening for cognitive impairment in a stroke population, only had a sensitivity score of .62 (specificity of 0.88), which is only slightly higher than chance (Blake et al., 2002).

The problems associated with the MMSE are further highlighted in the relationship between scores on the MMSE and on full scale neuropsychological test batteries using clinical samples. Nys et al. (2005a) conducted a study investigating the efficacy of the MMSE as a screening instrument for cognitive impairment in an acute stroke population. Thirty-eight patients diagnosed with an Ischaemic stroke or intra-cerebral hemorrhage were recruited within two weeks of their stroke ( $M = 6.5$  days,  $SD = 2.9$ ). Patients were assessed in a single session on the MMSE as well as a customized neuropsychological battery that assessed six domains of abstract reasoning, verbal memory, executive function, visual memory and language. Comparison diagnoses based on the neuropsychological battery and the MMSE indicated that the accuracy of the MMSE in detecting cognitive impairment was no better than chance with sensitivity of .67. Moreover, no optimum MMSE cut-off value

could be identified. The MMSE was particularly insensitive to impairments in abstract reasoning, executive functioning, and visual perception/construction. The Nys et al. (2005a) study confirms that while the MMSE might be useful for detecting an overall cognitive impairment, it is not overly useful in measuring impairments at the level of different cognitive domains.

**2.3.1.2 Montreal Cognitive Assessment.** The MoCA is a 30-point screen of cognitive functioning. It measures the six cognitive domains of short term memory, visuospatial functioning, executive functioning, attention, concentration and working memory, language and orientation. These domains are those typically used in neuropsychological test batteries and in many instances are simply abbreviated forms of standard neuropsychology tests. Memory is measured by the recollection of five nouns after a five minute time lapse. Visuo-spatial functioning is measured by two tasks, the first being the task of drawing a clock, the second is copying a drawing of a cube. Executive functioning is assessed by the abbreviated Trail Making B - a speeded task of drawing a trail between a combination of letters and numbers, a word fluency task - naming as many words as possible beginning with the letter "F", and a similarities task - describing the similarity between two objects such as a banana and orange. A target detection requiring the participant to tap their finger every time they hear an auditory cue is used to measure attention. Concentration and working memory are measured using a serial subtraction task (minus seven from 100 continuously) and abbreviated forwards and backwards digit span. Language is assessed by naming line drawings of three animals and a sentence repetition task. Lastly, orientation assesses the patient's understanding of the current time and place.

The MoCA is statistically superior to the MMSE, with higher reliability (Cronbach  $\alpha = .78$  for MoCA and  $\alpha = .60$ ), fewer issues of ceiling effects, and a higher sensitivity to cognitive impairment (sensitivity = 89% for MoCA and 63% for MMSE) (Nasreddine et al.,

2005). While some of the cognitive constructs assessed are common to both screening test, the MoCA incorporates more detailed and varied tasks. In particular, the MoCA screens thoroughly for visuospatial problems as well as executive functioning deficits, making it more suitable for detecting deficits common to patients with stroke.

### **2.3.2 Neuropsychological Tests**

The second phase of neuropsychological assessment post-stroke administering a battery of neuropsychological tests that covers a wide range of cognitive abilities. Such assessment can be via a standardised set of tasks such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) or can be a customised battery consisting of a compilation of relevant tests to suit particular circumstances. In most instances the tests that are used have well-documented psychometric properties and normative data which is used to determine the presence or absence of an impairment. In most instances an impairment is recognised when performance falls below a z-score of 1.96 below the average of the normative sample.

While detecting the presence of impairment is relatively straight forward, identifying the cognitive functions measured by the test batteries can be problematic. For a start the domain itself might be ill-defined. The area of executive functions has been and remains the domain that suffers in this respect. As Kramer et al. (2014) point out.

When executive functioning is targeted in research, there is considerable variability in how it is operationally defined. Tasks purportedly measuring fluency, working memory, concept formation, set shifting, inhibition, organization, abstract reasoning, and novel problem solving, either individually or in various combinations, are all used as markers of executive functioning, with the implicit assumption that these tasks measure the same construct. (p. 11)

The problem is exacerbated when multiple tests from multiple domains are administered in customised test batteries. Thus, tests are selected on the presumption that each tests measures a specific cognitive domain, but the underlying structure of the test battery is frequently not checked to confirm the assumed structure. In the instances where researchers have attempted to determine the structure of a customised battery they have done so statistically by using factor analysis (e.g., Planton et al., 2012) and or structural equation modelling (e.g., Miyake et al., 2000). The aim of these statistical techniques is to identify those tasks that share aspects in common, and at the same time differentiate them from other tasks. In the case of factor analysis, most analyses will produce more than one factor. Labels are then devised to identify the commonality among the tests that define a specific factor. Consequently, terms like executive functioning, speed of processing, memory, visuo-spatial processing, attention, and language represent the end result of many such studies. Table 2.1 depicts one example of a factor analysis conducted on data derived from a customised battery of tests that was administered to a group of patients with stroke (Planton et al., 2012).

One problem with factor analysis is that the factors that emerge are very much dependent on what tests were administered. A particular test might load on one factor in one customized test battery and load on a different factor on a battery consisting of alternative tests. This means that the one test can often be reported under different domain names. As an example, both forward and backward digit span are seen as measures of verbal working memory in the Planton et al. (2012) study. In other studies, digit span is seen as a measure of verbal memory (Nys et al., 2005a) or attention (MoCA; Folstein, Folstein, & McHugh, 1975) and yet in others forward span is seen as a measure of attention and backward span is seen as a measure of working memory (Lezak et al., 2012; Sachdev, Brodaty, Valenzuela, Lorentz, & Koschera, 2004a). Therefore, for many tests, there may be little consensus on what the particular test measures.

Table 2.1

*Cognitive Domains and the Tests Associated with those Domains*

Function	Test Used
<b>Executive Functions</b>	
Mental speed processing	Trail Making Test A; Stroop-colour; Stroop-read; Digit symbol
Motor speed Processing	TEA battery (divided attention), reaction; TEA battery (flexibility), reaction time; TEA battery (go-no-go) reaction time.
Verbal working memory	WAIS-R digit span, forward; WAIS-R digit span, backward
Visual working memory	WMS-R spatial span, forward; WMS-R spatial span, backward
Inhibition	Stroop interference, Stroop test, subtraction score
Initiation	Verbal fluency, fruit category; Verbal fluency, letter R
Flexibility	Trail Making Test, subtraction score; TEA battery (Flexibility)
Categorization	Modified Card Sorting Test
<b>Attention</b>	
Continuous attention	TEA battery (divided attention); TEA battery (flexibility); TEA battery (go-no-go),
<b>Memory</b>	
Free recall	RL/RI-16, sum of three free recall; RL/RI-16, delayed free recall; Rey memory, score
Cued recall	RL/RI-16, sum of three free recall; RL/RI-16, delayed free recall
Recognition	RL/RI-16, recognition score; DMS48, set 1 score

*Note.* TEA = Test of Everyday Attention; WMS = Wechsler Memory Scale; WAIS-R = Wechsler Adult Intelligence Scale-Revised; RL/RI-16 = free recall/cued recall-16 = a French equivalent of the Free and Cued Selective Recall test.

In addition, some deficits show up across domain tasks. The Planton et al. (2012) study showed that tasks like the Trail Making Test and the Stroop test have components that load on different factors, in this case mental processing speed, inhibition and flexibility, even though they are all speeded tasks. Yet it is not uncommon for cognitive deficits to appear across all speeded tasks, not just a subset of specific tests (Hochstenbach, Mulder, van Limbeek, Donders, & Schoonderwalt, 1998). Deficits in speed of processing may be reflected in deficits on specific tasks designed to measure that function, but they can also show up on

cross-domain tasks as well. The implications of this are that cognitive deficits can be detected in tasks that are deemed to measure a specific cognitive domain, or through deficits across a number of tasks from different domains. Both individual task performance and a pattern of deficits across tasks need to be considered.

While specific neuropsychological tests can be contentious in what they measure, it remains the case that neuropsychological test batteries have become the norm in assessing cognitive impairment. Most of the individual tests have well-validated psychometric properties that allow for relatively accurate detection of impairment. A negative of neuropsychological test batteries is that while a deficit on a test can readily be detected, the nature of that deficit is often open to interpretation, depending on how a specific task is categorised by the authors. This imprecision in the domain-test linkage has an impact on cross-study comparisons exploring the cognitive domains affected by stroke. Deficits on the same set of tests, say forward and backward digit span, may be reported as a deficit in one particular domain in one study and a different domain in another study. Consequently, the ability to identify a consistent pattern in the frequency with which different domains are impaired is likely to be severely compromised. Thus, in reviewing the cognitive domains affected by stroke, it is not expected that a totally consistent picture will emerge.

#### **2.4 Cognitive Domains affected by Stroke**

When exploring the cognitive effects of stroke, the literature is limited to those studies that have utilised a cognitive battery that addresses the neuropsychological dimensions described above in Section 2.2. The identified studies are organised using the acute/chronic framework adopted in Chapter 1 and are presented in chronological order. The intent is to simply describe each study and to draw conclusions after the relevant studies have all been reviewed.

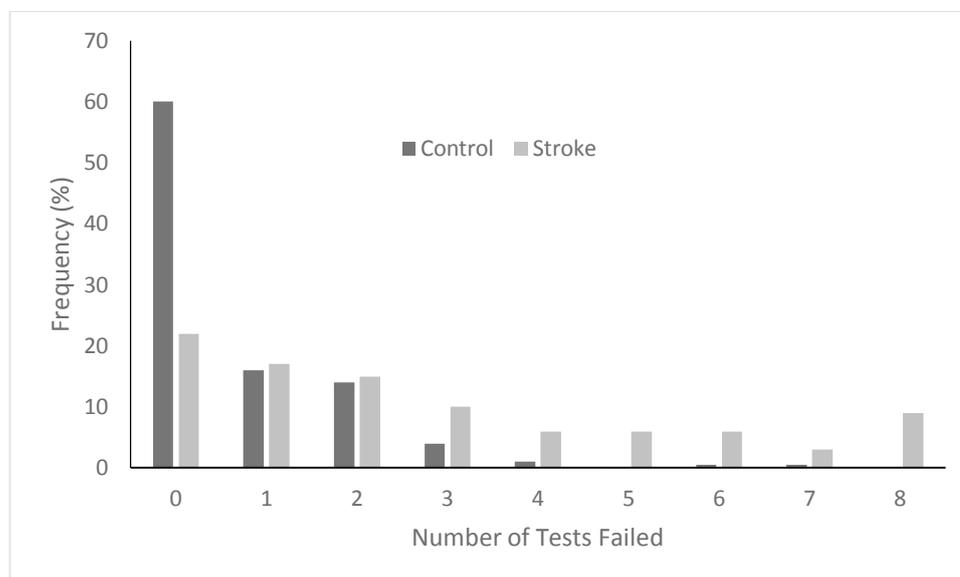
Three sources of information were sought in reviewing each study. The first was the number of people classified with at least one identifiable cognitive deficit. If this measure is less than 100% it means that some people can have a stroke and not suffer any cognitive consequences. The second measure deals with the degree to which people suffered a single impairment or multiple impairments. This is used to gauge the severity of the stroke. The more severe the stroke, the wider the range of impairments (Lezak et al., 2012). The third measure dealt with the cognitive domains that were most frequently affected.

#### **2.4.1 Acute Phase**

There are few studies that have explored testing during the early stages post-stroke. In fact, many studies purposely do not test during this period and wait some months until behaviour stabilises. Moreover, in the acute phase many people are not able to do the tests because of illness, increased need for sleep, and inability to maintain concentration, and these are sources of exclusion from the study (Jaillard, Naegele, Trabucco-Miguel, LeBas, & Hommel, 2009; Nys, van Zandvoort, de Kort, Jansen, de Haan, & Kappelle, 2007). In spite of these difficulties, there are a small number of studies that have used batteries of neuropsychological tests and reported how many patients with stroke show a cognitive deficit in at least one domain, how many patients show multiple deficits in multiple domains and which domains are most commonly affected by stroke.

An early study by Tatemichi, Desmond, Stern, Paik, Sano and Bagiella (1994) collected neuropsychological data across a 3-year period from 251 ischaemic patients with stroke who were over 60 years of age. Of this sample, 227 were tested within 30 days of the stroke. Their performance on 17 tests that explored memory, orientation, verbal skills, visuo-spatial ability, attention, and abstract reasoning was compared to 240 community-living age-matched controls. Participants were considered to have failed a test if they were outside the

95% confidence interval defined by normative performance (the control group). Figure 2.1 shows the failure rates for the two groups.



*Figure 2.1.* Frequency of test failure between control and stroke groups in Tatemichi et al. (1994).

As Figure 2.1 indicates, in the acute stage post-patients with stroke tended to fail more tests ( $M = 3.4$ ,  $SD = 3.6$ ) than controls ( $M = .08$ ,  $SD = 1.3$ ), with approximately 78% of the patients with stroke having an impairment in at least one domain with 40% of the control group also having an impairment in one domain. The distribution of failed tests also differed between the groups. The patients with stroke were more likely to be impaired across multiple domains than the control group. Individual differences are also clearly evident in the figure. There were a substantial number of patients with stroke who failed only one or two tests, and there were those who failed multiple tests. The implication of these data is that strokes can either have selective effects, in that one or two tests might show impairment while other tests show normal functioning, or alternatively, stroke can have a widespread impact on cognitive functioning. The importance of individual difference results is reinforced by their finding that in the stroke group there was a clear linkage between the number of tests failed and

functional impairment. That is, those who experienced deficits across multiple domains experienced more problems in daily living.

When looking at group differences, Tatemichi et al. (1994) found that on all tests the patients with strokes scores were significantly below those of the control group. A logistic regression was conducted to see which tests and which domains best discriminated between the groups. The three tests that discriminated between groups were a speeded detection task (38.5% failure rate), semantic fluency (32.7% failure rate; note phonemic fluency failure rate was 14.4%) and memory recognition test (24.6% failure rate). Consistent with these test results, attention, memory, language and orientation were the domains that best discriminated between the groups, although there were still differences in visuo-spatial and abstract reasoning domains.

Nys et al., (2007) also investigated deficits in the acute phase. Data for this research was collected over a period of 2 years across three hospitals with the data set consisting of 228 participants. Refusals to participate and the application of various exclusion criteria (death, age, recovery within 24 hours (TIA), pre-existing cognitive deficit, alcohol dependence, etc) reduced the final sample to 190 who had suffered a stroke within the previous 3 weeks (Mean = 7.9 days). The customised test battery involved 15 tests that assessed abstract reasoning, verbal memory, executive functioning, visual perception and construction, visual memory, language and neglect. In this study the criterion for a deficit was deemed to be 1.67 standard deviations from the normative mean on each test. With this criterion, cognitive impairment in at least one domain was found in 74% of the patients with stroke. The distribution of failed tests was very similar to that found by Tatemachi et al. (1994) (see Figure 2.1) where the likelihood of multiple domains being affected was more prevalent with haemorrhagic strokes than ischaemic strokes. The most prevalent failures were in executive functioning (39.1%) followed by visual perception and construction (38.1%),

neglect (31.3%), abstract reasoning, verbal memory, language (all 25%), and visual memory (22%). The study confirms the widespread impact of stroke, the different levels of severity among the patients, and the diverse range of domains that were affected.

Over the course of 4.5 years, Jaillard et al., (2009) examined 879 patients who had suffered their first ischaemic stroke. Of these, 668 were excluded because their performance on the MMSE indicated that they did not have a cognitive impairment. Thirty-four were excluded because they could not do the tests at all, they could not be tested within the required timeframe, or because they refused to participate. Performance on the final sample of 177 patients was compared to 81 healthy control participants who were matched for age and education. Approximately 2 weeks after their stroke, participants were administered a customised battery of 23 tests deemed to assess the domains of short-term and episodic memory, executive functions, working memory, and general and instrumental functions. Performance of the control group defined normative performance, and a cut-off of between 1 and 2 standard deviations below the norm was used to define mild impairments, and a cut-off of 2 or greater standard deviations defined moderate impairments. Patients performed significantly below the control group on all tests except five. Using the more lenient criteria of 1 standard deviation below the mean, 91.5% of participants showed a cognitive deficit in at least one cognitive domain (which is not surprising given that they were selected on the basis of the MMSE indicating that they had a cognitive deficit). Using the more stringent criteria of a z-score of -2 or greater, 73.4% of participants showed a cognitive deficit in at least one area. The distribution of the single versus multiple domains affected was not reported. The domains most affected were working memory (87.6%), executive functioning (64.4%) and short-term and episodic memory (64.4%), with general and instrumental functions being less impaired (24.9%). Given the selective nature of the sample in the study it

was not surprising that prevalence rates were much higher than the other studies. However, problems in executive functions and working memory were still frequently detected.

Hurford, Charidimou, Fox, Cipolitti and Werring (2013) considered a population of 1,335 people admitted to a hospital in London over the course of 7 years. The principle exclusion criteria in this study were lack of imaging data, lack of neuropsychological data, and whether the stroke had been resolved within 24 hours (TIAs). The final 209 participants were administered a customised battery that measured six cognitive domains: verbal memory, visual memory, naming, perceptual functions, executive function, and processing speed and attention functions. Patients were only tested once at different stages during the recovery process. They were tested either less than 1 month after their stroke (acute phase), between 1 month and 3 months (chronic stage) or after 3 months (chronic phase) following the stroke. A cognitive deficit was determined by either a 5<sup>th</sup> percentile cut-off or outside a 95% confidence interval, depending on the particular test. When testing took place during the acute phase, they found that 72.4% of the 129 patients showed cognitive deficits in at least one domain. They did not present data showing the distribution of multiple deficits, but there were clear differences in which domains were most affected. The deficits were most pronounced in speed and attention (70% failure), executive functioning (35% failure), with naming, verbal memory and perceptual skills showing similar failure rates (around 30%) and visual memory being the least affected (15%).

The four studies that have involved neuropsychological testing during the acute phase following stroke converge on a number of findings. Firstly, very few people (20% to 30%) who have a stroke escape without some form of cognitive impairment. Secondly, the number of domains that are affected following stroke vary from person to person. Some people exhibit impairment in one or two domains, but others experience difficulties across multiple domains. Moreover, although not readily evident in the data, it can be surmised that there is

no single profile of cognitive impairment. For example, in the Tatemichi (1994) study, 10% of the patients with stroke showed cognitive impairments in three areas. While the specific areas were not described, it is safe to conclude that the three areas could vary from person to person. Thirdly, all the cognitive domains can be affected by stroke. Processing speed, attention, working memory, and executive functioning appear to be the domains that are most widely affected, but other domains are still susceptible to impairment.

#### **2.4.2 Chronic Phase**

There have been more studies examining the prevalence of cognitive impairment where testing has occurred 3 months or later post-stroke. While the acute phase studies were all conducted in a hospital setting, when looking at studies in the chronic phase some are conducted in a hospital setting and some involve patients who have been discharged from hospital. In the case of one study, participants were selected on the basis that they had made an excellent functional recovery. Thus, there is a much more diverse range of patients and testing conditions than was the case in the studies involving testing in the first few weeks post-stroke.

In one of the earliest large studies concerned with prevalence rates (Hochstenbach, Mulder, van Limbeek, Donders, & Schoonderwalt, 1998), the cognitive domains of orientation, memory, attention, visuo-spatial/construction, language and arithmetic were compared between 229 patients with stroke and 33 controls. The patients with stroke were recruited over 3 years from three different hospitals and it is not clear what the original sample was or what inclusion/exclusion criteria were applied. The patients with stroke were tested in the hospital with testing occurring on an average of 72 days post-stroke. Those who had suffered a stroke showed significantly poorer performance than the control group on most of the tasks and in all domains. A cut-off of 2 standard deviations below normative performance was used to establish a cognitive deficit. The distribution of zero, single or

multiple deficits was not published, instead the percentage of the participants showing a deficit for each task was presented. It was clear that the frequency of impairments within a domain varied greatly. With the memory domain, for example, 40% of participants showed a deficit on the backward digit span test, but only 4% of participants showed impairment on the WAIS-information subtest. Likewise, for the attention domain, 56% showed a clinical impairment on Trail Making Test-B but only 10% showed a deficit on a picture scanning task. Although speed of processing was not explicitly tested, many of the tests used had a time pressure component. The results indicated that there were widespread failures (between 40% and 50%) in any task that had a time component. The results of this study confirmed that cognitive deficits due to stroke could be found on most cognitive tests with some being more sensitive than others. In this particular data set, speed of processing appeared problematic with specific tasks such as backward digit span and verbal fluency tasks having high prevalence rates.

Madureira, Guerreiro and Ferro's (2001) study examined the course of cognitive deficits in haemorrhagic and ischaemic strokes at 3 months post-stroke. The data was collected over a 2-year period with the final sample in the study consisting of 237 patients. The starting sample, the number who were excluded for any reason and the number of refusals and dropouts were not reported. Using a comprehensive customised neuropsychological test battery, patients were tested on various cognitive domains including memory, verbal fluency, language, visuo-constructional abilities and abstract reasoning. Of the 237 patients assessed, 55.3% showed an impaired performance on cognitive screening tests and were assessed further. Fifty percent of fully assessed patients showed cognitive deficits. The overall prevalence rate for having a cognitive impairment was 27.7%. Again the distribution of impairments was not presented. Patients with stroke in this study were most

frequently impaired on tests of verbal fluency, orientation, logical memory and associative learning.

The importance of extraneous factors in identifying the level of impairment is reflected in another study in which a selection of neuropsychological tests were administered to patients with stroke three months after the stroke. Srikanth et al. (2003) had referrals from 23 public hospitals, 36 private hospitals, general medical practitioners and residential care facilities over a 13-month period. Of the 458 referrals, death and exclusion criteria reduced the potential sample to 148, and of these 99 people consented to the research. The final sample involved 99 community living participants who were tested on a customised battery measuring attention, orientation, memory, spatial ability, language, and executive ability. Performance was compared to a matched control group. They found that for their patients with stroke, cognitive impairment was more likely to occur in a single domain than across multiple domains. They did not report prevalence data but they did report test failure data. Failures were most apparent in memory (26%) and spatial processing (26%), less apparent in executive functioning, language and attention (all = 15%) and least observed in orientation (10%). What was surprising in this study was that the control group showed similar patterns of test failure. Compared to the studies conducted while the patient was in hospital, the community-based sample in this study exhibited substantially less impairment.

Sachdev et al. (2004a) recruited participants through two large hospitals across a 3-year period. Two hundred and fifty-two of a total sample of 1,050 (550 did not consent to participate) were considered eligible, but only 170 patients and 96 age-matched controls consented to participate. Patients and controls were assessed using a customised neuropsychological battery between three and six months after a stroke and again approximately a year later ( $M= 14.6$  months,  $SD= 3.5$  months). The battery assessed verbal memory, visual memory, working memory, attention, mental control, language, information

processing speed, visuo-construction, abstract reasoning and executive functioning. In comparison to controls, patients with stroke showed a significantly poorer cognitive performance in the areas of executive function, processing speed, visual memory, abstract reasoning and visuo-construction, thus replicating many of the studies reviewed earlier.

With regards to prevalence data, the control group provided the normative data on each of the tests and impairment was determined by using the 5<sup>th</sup> percentile criterion. With this criterion, 41% of the sample was cognitively intact, 36% had a cognitive impairment in one domain but did not show evidence of dementia as measured by the DSM-IV, and 23% showed evidence of dementia. In comparing the cognitively intact with the cognitively impaired without dementia, the greatest discrepancies were in the domains of executive functions, language, abstract reasoning and speed of processing.

Planton et al. (2012) conducted a study where the aim was to examine the cognitive performance of people who had suffered a stroke in the prior 3 months and were judged to have made a complete recovery in terms of motor, linguistic and functional outcomes. The data were collected over a 14-month period from patients who attended a stroke clinic. Of the 308 patients who were screened, 199 were excluded because of death, age, and failures on the critical inclusion criterion of no impairments in the areas of motor, linguistic and functional behaviour. The remaining 109 were pre-recruited, 98 were assessed as outpatients (11 lost in follow-up) with 38 being excluded because of emerging issues between recruitment and assessment. Sixty patients who were recruited for the study were administered the test battery, outlined in Table 2.1, 109 days on average after their stroke. This battery consisted of 13 functions, eight of which measured executive functioning, two of which measured attention and three which measured memory. The patient data were compared to 40 age and education matched controls. In this study scores that were two standard deviations below normative performance defined a cognitive deficit. They found that 40% of the stroke group

had a cognitive deficit in at least one domain. When comparing test performance between the stroke and control groups, the patients with stroke produced significantly lower test scores on all domains, with the largest deficits being found in executive functioning (categorisation, mental processing speed, initiation, verbal working memory). There were also substantial deficits in attention (continuous attention) and recall (free recall).

The above studies showed much the same pattern as the studies involving testing during the acute phase. That is, cognitive deficits are quite pervasive following the first few months post-stroke, even when patients have made a full functional recovery. Deficits can be found in most cognitive domains, but problems in executive functioning and speeded tests appear to be particularly vulnerable. The literature indicates that the main difference in testing outcomes between the chronic and acute phase is that prevalence and test failure rates appear to be much higher in the studies where testing is done in the acute phase (70%-80%) than when testing is done in the chronic phase (20%-40%). This outcome suggests that natural recovery occurs in the cognitive domain just as it does in the motor domain.

### **2.4.3 Cautions**

While the above picture suggests that there is a very high prevalence of cognitive impairments following stroke, there are two factors that suggest the possibility that there might be over-estimations of impairment. The Tatemichi (1994) study shows that a portion of the control group also failed on one or more of the neuropsychological tests. This raises the possibility that some of the failures in the stroke group might well be normal age related deficits rather than being a direct result of the stroke. In addition, none of the studies have had co-morbid diseases as a stated exclusion criterion. This issue will be addressed in more detail in Chapter 6.

The second deals with subtle motor or speech effects. In the research described above, participants were only excluded if there were obvious and severe impairment that would

prevent the participant from completing the formal assessment tasks. It is quite possible that subtle differences in motor or speech problems could result in impaired performance that are not necessarily reflective of cognitive problem. Speeded motor or verbal tasks, would be particularly prone to these subtle influences. While subtle speech and motor problems might over-estimate prevalence rates, the Planton et al. (2012) study of patients with stroke who had no motor, speech or sensory issues, indicated that strong cognitive impairments were readily apparent among the stroke group.

## **2.5 Natural Recovery of Cognitive Functions**

There is a small body of research that explicitly explores natural recovery of cognitive functions in the weeks and months post-stroke. The process of natural recovery from acute phase to chronic phase is important to explore as it provides a guide for potential benchmarks that the clinician can use to manage patient and family expectations about the future.

Evaluating the process of recovery is problematic in studies that involve a simple examination of initial versus follow-up test scores without the presence of a control group. In the case where performance deteriorates from initial testing, the effects of stroke could become more pronounced over time, or alternatively this deterioration simply reflects normal aging processes given that normal ageing is associated with decrements in cognitive performance. An elderly person who suffers a stroke may show decreased levels in cognitive function when tested a year later, but those effects might be due to normal ageing processes rather than a direct result of having suffered a stroke (Anstey & Low, 2004). Alternative explanations are also possible in the case where performance improves over time. There may well be recovery in the nominated cognitive domain, there is just increased familiarity with doing the cognitive tests, or there is regression to the mean. Those tests that employ a control

group probably give a clearer indication of the recovery process than do those that simply explore initial and follow-up test results.

### **2.5.1 Recovery in the Chronic Phase**

In Chapter 1 the recovery process during the chronic phase was characterised by slow and gradual improvements in performance. That literature did not, however, provide any information about the longevity of any improvement or acknowledge the possibility that improvement at one point in time could be followed up by decrements at another point in time. The following studies address these issues. Two studies use the MMSE as the measurement tool where global cognitive impairment is the focus of the studies. Other studies use a neuropsychological test battery that allows for the assessment in different domains. Again the studies will be presented in chronological order with integration of the studies being addressed when all studies have been reviewed.

**2.5.1.1 MSSE.** In one of the longest studies on the cognitive recovery from stroke conducted by Patel, Coshall, Rudd and Wolfe (2003), cognitive performance was first measured at three months and then again at 1-year, 2-years, and 3-years post-stroke using the MMSE as the assessment tool. The initial study obtained permission from 294 participants with first-ever strokes. A third of the participants had died prior to testing at the 3-month point and others were excluded because they were too dysphasic to undergo the cognitive tests. In the end, 163 participants were tested 3 months after the stroke. At this point, 39% of the sample displayed cognitive impairment as measured by the MMSE. Thus, the prevalence data in this study is in the same ballpark as the prevalence rates reported in the studies in Section 2.4.2 above. At the 1-year mark 35% were impaired; after 2 years 30% were impaired, and after 3 years 32% were impaired. These results confirm that changes during the chronic phase are not large with change occurring gradually.

There were two other aspects of the data that are worth commenting on. Those who were initially diagnosed as impaired were more likely to die during the 3-year period than those who did not have an initial impairment. The death rates at 1 year were 23% for the impaired and 8% for the intact groups. At the 2-year point these figures were 36% and 15%, and at the 3-year mark the figures were 45% and 24%. In short, the level of impairment at initial testing were positively correlated with mortality rates.

Secondly, for those that did not die during the study period, the trajectory looks quite similar to the mortality data. The development of those who were originally classified as having no impairment (99 people) and those that were originally classified as having an impairment (64 people) was traced over the 3 year period. Of the intact participants, 74 of the 85 living participants were still intact, and 11 had deteriorated at the 1-year point. Of those 74 who were intact at the 1-year mark, 50 were intact at 3 years, 12 had died and seven were impaired (and five dropped out). Twenty people who were intact at 3 months went on to develop problems at subsequent test points. Of these 20, seven recovered at a later stage, with only six participants alive and impaired at the 3-year mark.

In a similar study exploring recovery using the MMSE as the measurement tool Liman, Heuschmann, Endres, Floel, Schwab and Kolominsky-Rabas (2011) collected data across a two year period from multiple sources in a large European city. The 1,631 patients considered in the study were all first-time patients with stroke. Of these patients, 893 were considered eligible for the stud, and 678 were actually tested. The data presented in the study consisted of the 630 participants who were tested 3 months after their stroke. Consenting participants from the original set were tested again at 12 months and at 3 years.

At initial testing 14.8% of the group were considered to have a cognitive impairment (MMSE score <24) while 85.2% considered cognitively intact. Of the cognitive intact group 85% remained cognitively intact and 5% developed an impairment at the 1-year mark. Of the

intact participants at the 1-year mark, 83% remained intact at the 3-year point and 4% had acquired an impairment. Thus, the vast majority of participants who had no cognitive impairment at the 3-month point did not develop impairments across the next 3 years.

Twenty five percent of the people who were initially diagnosed as having an impairment recovered by the 1-year point, and 56% of those maintained their cognitive status at the 3-year mark. Forty five percent of the impaired group were still impaired at the 1-year point, and 56% of those were still impaired at the 3-year point but 10% had recovered at the 3-year mark. The two groups differed on mortality rates at the 1-year mark; the mortality rate was 15% for the cognitively intact group, and 32% for the impaired group.

Results from the two studies show the same pattern of effects. Firstly, the prognosis for those with intact cognition at the 3-month point was quite positive with subsequent declines in cognition possibly related to age. The prognosis for those who had impairments at 3-months was not as positive. Such participants were more likely to die, or to remain impaired as a result of the stroke, advancing age or both. There were very few participants who recovered after three or 12 months and maintained their intact status over the remaining years.

**2.5.1.2 Neuropsychological assessment.** In perhaps the first study to explore recovery after stroke, Desmond, Moroney, Sano and Stern (1996) administered a battery of neuropsychological tests to 151 patients with stroke at 3 months and then annually after that for 3 years. The battery included measures of verbal and nonverbal memory, orientation, language, visuo-spatial functioning, verbal and nonverbal abstract reasoning skills, and attention. They did not examine performance changes at the domain level. Instead they created a summary score by first transforming their test results into  $z$  scores (based on the performance of 192 normative controls) and then averaging those domain scores.

Improvement was defined as an increase in that summary score greater than two standard deviations above the mean change in the normative group. That is, improvement had to be substantially above any improvement that the control group produced. They estimated that of their stroke sample 35% showed a cognitive impairment. Of these, 35.9% (12.6% of all patients with stroke) produced the criterion level of improvement. This improvement was strongest at the 1-year test, with no further improvements being noted in subsequent years. Those who showed the most improvement were those who showed the greatest deficits at baseline with marked improvements in memory, orientation, visuospatial function, and attention. There were no significant improvements in language or abstract reasoning.

Tham et al. (2002) explored recovery rates in a sample of Singaporean patients with stroke who were first tested within 6 months of their stroke and then tested 12 months later. Over a 2 year period, 270 patients with stroke were invited to participate in the study with 252 consenting to do so. The patients were administered the Vascular Dementia Battery which consists of 24 tests that assess attention, language, visual memory, verbal memory, visuo-spatial construction and visual speed of processing. The criteria used to determine impairment were not reported. At baseline, 56% of patients were cognitively intact, 40% were cognitively impaired in at least one domain but not demented according to DSM-IV diagnosis, and 4% were diagnosed with dementia. Of the 155 patients who were reassessed 1 year later, 77% had the same classification as at baseline. All those who were demented at baseline remained demented at follow-up. Ten percent of those originally classified as cognitively intact were subsequently classified as having an impairment, but none were demented. Of the patients classified as having an impairment at baseline, 30.6 % were reclassified as having no impairment (recovery had taken place) and 10% had deteriorated and had dementia. When comparing those who remained stable to those who had deteriorated, the differences were found in only 8 of the 24 tasks.

Ballard, Rowan, Stephens, Kalaria and Kenny (2003) likewise tested a group of 115 stroke sufferers over the age of 75 (no sampling data reported) at 3 months and then again at 15 months on the MMSE and Section B of the Cambridge Assessment of Mental Disorders for the Elderly (CAMCOG). The CAMCOG provides measure of global cognitive functioning as well as scores for memory, orientation, language comprehension, language expression, attention, praxis, calculation, abstract thinking, perception, and executive function. They divided their participants into three groups: those who showed clinical signs of dementia at 15 months (9%); those who showed a positive improvement of 2 points or more on the MMSE (16%); and those who remained stable but not having dementia and large positive changes in MMSE scores (75%). However, 30% of the stable group had scores that actually declined over the 15-month period. In recalibrating their groups to include a deterioration of any kind, be it dementia or negative change on the MMSE, 31% of the sample showed deterioration, 16% showed marked improvement, and 53% showed little change at all. When performance on the test battery was examined, the stable group showed statistically significant improvements in abstract reasoning, executive functioning, and verbal fluency in the stable group. Those in the improved group also showed statistically significant improvement in abstract reasoning and executive functioning, but also in language, memory, and attention.

In a similar study, Hochstenbach, den Otten and Mulder (2003) administered a customised test battery to 65 people (from a total sample of 92, of whom 27 declined to participate) who had suffered a stroke and to 33 controls. The battery assessed the domains of orientation, memory, attention, language, and visuo-spatial processing. The initial test occurred at 2.7 months on average and the second follow-up testing session was completed at 27.7 months on average. Scores on all domains had significantly improved at the second testing. The authors also explored the extent to which all participants improved. They divided

their sample into those who showed improvements of greater than 2 standard deviations above any improvement displayed by the control group. Collapsed across the six domains, 76% of the participants remained stable, 17.5% showed improvement, and 5.9% deteriorated (less than 2 standard deviations below controls). This change was not equally evident among all domains. For example, in the memory domain 93.5% of the group remained stable, 6% improved and .5% deteriorated. In contrast, in the attentional domain 27.7% showed improvement and 59.4% remained stable, however 12.9% deteriorated over the 2-year period. The other domains were intermediate with lower rates of both improvement and deterioration.

Sachdev et al. (2004b) did a follow-up study based on the Sachdev et al. (2004a) data reported in Section 2.4.2 above, where the recruitment and participant selection details were described. The 170 patients and 96 controls were assessed using a customised neuropsychological battery between three and six months after a stroke and again approximately a year later ( $M= 14.6$  months,  $SD= 3.5$  months). The battery contained tests that assessed verbal memory, visual memory, attention, mental control, language, information processing speed, visuo-construction, abstract reasoning and executive functioning.

The focus in this follow-up study was on cognitive decline rather than on recovery, moreover, in the original study (Sachdev et al., 2004a) the patients with stroke were divided into those with intact cognition and those who were impaired. In the current study these groups were collapsed into one group, so it was not possible to track the changes in cognition for the two distinct groups. In the initial test, patients with stroke (as a whole) showed a significantly poorer cognitive performance in the areas of executive function, processing speed, visual memory, abstract reasoning and visuo-construction on the first testing occasion, thus replicating many of the studies reviewed earlier. After 1 year, patients were assessed again to investigate the change in cognitive status over time. Results indicated that the

cognitive domain scores did not improve over the 1 year period, with further deterioration in the areas of verbal memory and visuo-constructional abilities. The one exception to this pattern was the improvement in visual memory performance.

The studies reviewed above show some commonalities and some differences. The first commonality is that by and large cognitive performance remains relatively stable between 3 months and 3 years post-stroke. Gains can be detected but the gains are relatively modest. Very few people make substantial recovery during this period. It is also true that some people deteriorate during this period, but it seems that more improvement than deterioration occurs. Some of the differences occur in which domains show improvement and which domains do not show improvement. It appears that executive functioning seems to be consistently implicated as an area that is both susceptible to stroke and strong recovery. Attention and abstract reasoning also show similar changes. As a final point it is worth noting that those studies that report prevalence data suggest that when first tested, between 20% and 40% of stroke victims show cognitive impairment. These values are in the same range as those studies described in Section 2.4.2 that have explicitly explored prevalence data.

### **2.5.2 Recovery in the Acute Phase**

While recovery in the acute phase of stroke has been studied in terms of motor problems, aphasia and neglect, there are just two studies that have looked at recovery involving the cognitive domains. One study involved repeated testing, with the first testing taking place within the first two weeks post-stroke. The second study involved a cross sectional design where different cohorts were tested at less than one month, between one and three months, and more than three months post-stroke.

Nys et al. (2005b) conducted a study where data was collected across three hospitals over a 2 year period. From a starting sample of 228 patients, 168 people who had experienced their first stroke and met inclusion criteria were administered a battery of neuropsychological

tests in the days following stroke (Mean = 7.9 days, SD = 4.2). Of the initial sample, 111 were tested again between three months and 10 months later (M = 7.5 months, SD = 1.3). A control group of 77 healthy controls were tested under the same conditions.

The battery involved tests that measured the domains of reasoning, language, visual memory, verbal memory, executive functioning and visuo-spatial processing. Domain scores were calculated by averaging the z-scores of the within-domain tests. A score of 1.67 standard deviations below the mean on any domain was classified as impaired performance in that domain. A person was classified as having a cognitive impairment if they met the criterion for impairment in any one of the six domains. Of the 111 patients, 49% were cognitively impaired in one or more domains in the first weeks after stroke, with a mean of 3 impaired domains per patient [range: 1–6]. The domains that were impaired differed in frequency with 32.4% of patients demonstrating impairment in visuo-spatial processing, 31.5% in executive functioning, 24.3% in abstract reasoning, 21.6% in language, 21.6% in verbal memory, and 16.2% in visual memory.

In the follow up test, patients with a small number of cognitive impairments at baseline more often demonstrated a complete cognitive recovery in the long term than patients with a more widespread cognitive impairment. Overall, the mean number of cognitive deficits per patient decreased from 3.0 at baseline to 1.3 at follow-up. Domain-specific recovery in visuo-spatial processing (83%) and visual memory (78%) were the most frequent, whereas recovery in abstract reasoning (41%) and language (54%) was the least common. A small number of patients showed deterioration characterised by acquiring a deficit in an additional domain during the intervening period. Of the patients with stroke who were intact at baseline, 90% remained unimpaired at follow-up.

When comparing test performance between the control, initially intact and initially impaired groups, the intact and control groups showed equivalent longitudinal effects. The

impaired group however, showed significantly improved performance in all domains with the largest gains found in visuo-spatial processing, verbal memory and executive functions.

Hurford et al. (2013) used a cross-sectional design to explore recovery in the domains of visual and verbal memory, visuo-spatial processing, executive functioning, language, and attention/processing speed. The sample characteristics were described in section 2.4.1. The test battery was administered to 209 people who had suffered ischaemic stroke. Participants were divided into three groups on the basis of when they were tested. The majority of participants (129) were tested less than 1 month after their stroke, 43 were tested between 1 month and 3 months, and 37 were tested after 3 months post-stroke. Impairment was calculated in different ways for different tests and different domains, but the comparative effects across different test periods are presented in Figure 2.2.

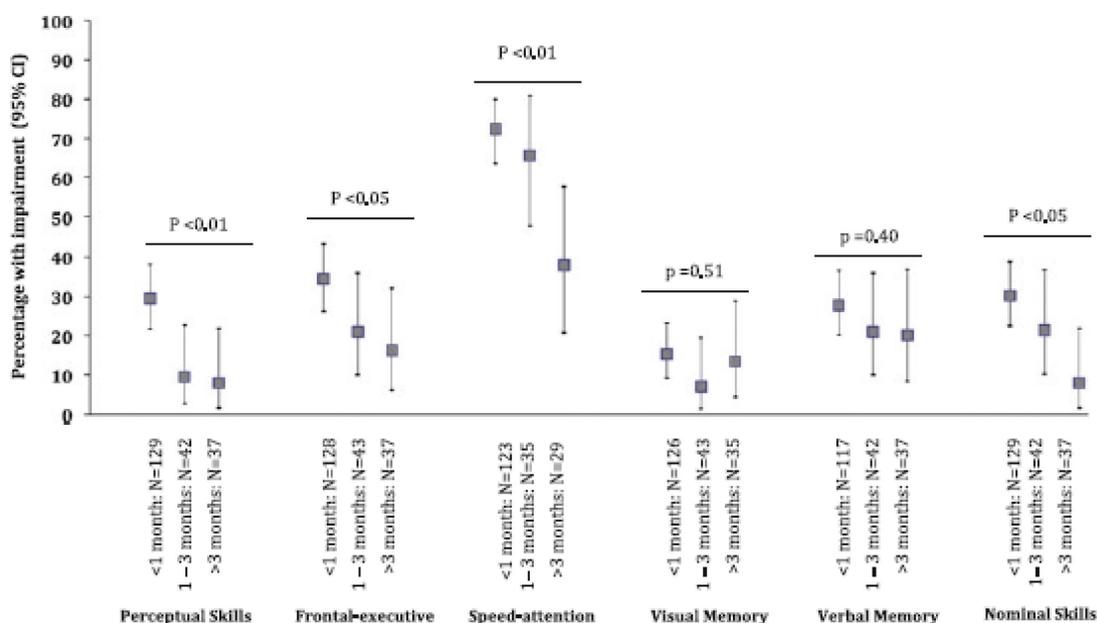


Figure 2.2. Prevalence of cognitive impairment at three stages in cognitive recovery. (Figure reproduced from Hurford et al., 2013).

The data in Figure 2.2 make a number of points. Firstly, impairment is more prevalent in the early weeks post-stroke than it is in later periods, suggesting that there is substantial recovery occurring during this time. Secondly, rates of impairment improve in the 1-month to 3-month period, and also in the greater than 3-month period, but the improvement in the latter period is less dramatic. Thirdly, recovery is more pronounced in some domains than it is in others. The speed and attention domain was the most impaired both acutely and after 3 months post-stroke, but this area produced the greatest decrease in the prevalence of impairment. Executive function, naming and visuo-spatial skills, although less prevalent overall, showed noticeable improvement between early and later testing. Visual and verbal memory demonstrated no significant trend for differences in the prevalence of impairment over time.

The studies that explored recovery when initial testing was done during the chronic phase indicated that changes during this period were not all that large. The studies in the acute phase showed the exact opposite. Improvements in the level of cognitive impairment were widespread and they were substantial.

### **2.5.3 Section Summary**

The previous sections have shown that cognitive deficits following stroke can be readily detected using neuropsychological test batteries. When exploring prevalence data on cognitive impairments following stroke, the studies in Section 2.4 consistently indicated that between 70% and 80% of patients with stroke were cognitively impaired in at least one cognitive domain when testing was conducted in the first few weeks following the stroke. However, when initial cognitive testing was conducted during the chronic phase, only 40% of patients showed cognitive impairments. The implication of this large change in prevalence rates is that cognitive recovery has taken place.

There are some studies that have explicitly explored recovery and those studies lead to the same conclusions as the prevalence studies. In the Nys et al. (2005b) study where initial testing was conducted during the acute phase, 70% of participants had a cognitive impairment with many participants having impairments across multiple domains. Recovery was observed approximately one year later in that the number of people showing a cognitive impairment declined, the number of people showing multiple impairments had declined, and in some areas there were no longer any participants showing abnormal behaviour.

The study by Hurford et al. (2013) led to very similar conclusions, although this data set suggests that recovery appears to be maximal over the first four weeks after a stroke. The implication of this data set is that recovery is a rapid process. However, the Hurford et al. study is a cohort study rather than a longitudinal study, so it remains to be seen whether the substantial changes observed in impairment rates between less than one month and between one and three months is truly reflective of the natural recovery process.

The Nys et al. (2005b) and the Hurford et al. (2013) studies track the recovery process from the acute phase to the chronic phase. Searches of the literature have failed to produce studies that have tracked the recovery process during the first twelve weeks post-stroke where performance has been assessed multiple times (at least twice) via a battery of tests that cover a range of neuropsychological domains. However, the acute to chronic recovery data suggests that if recovery is going to occur, improved task performance should be readily apparent during the acute period. One of the first aims of the current thesis is to address this gap in the literature by exploring changes in task performance associated with repeated neuropsychological testing during the acute phase of stroke.

A different picture emerged when initial testing was conducted during the chronic phase. In these studies impairment rates were generally consistent with the 20% to 40% level that had been observed in the prevalence studies (Hochstenbach et al., 1998; Madureira et al.,

2001; Planton et al., 2012; Sachdev et al., 2004a). In subsequent testing there was very little evidence for change. People who had no cognitive impairment at baseline continued to be cognitively intact at subsequent testing periods. Likewise, if there was a cognitive impairment at baseline, a small percentage of participants showed recovery at later stages but most participants were still impaired.

Evidence from the prevalence studies and the recovery studies converge on the conclusion that stroke can have a widespread deleterious effect on cognitive performance. However, improvement can be observed in the weeks, months and years after stroke. If recovery does occur, it is most pronounced in the weeks immediately following the stroke. Approximately 3 months later, behaviour stabilises such that subsequent gains are more modest.

When it comes to which aspects of cognition are most prone to impairment and which are most susceptible to recovery, the data were not as consistent. This was due in part to the differences in test batteries that were administered, the domains that were tested, the assignment of tests to cognitive domains, and the criteria used to classify cognitive impairments. In spite of this, there were some domains that occurred frequently. Executive functioning, speed of processing, visuo-spatial processing, some aspects of language (particularly verbal fluency), and working memory all appeared to be implicated in both impairment and recovery.

## **2.6 Cognitive Training**

The previous sections have indicated that natural recovery of cognitive functioning does occur (Hochstenbach et al., 1998; Madureira et al., 2001; Planton et al., 2012; Sachdev et al., 2004a), and that improvements are most pronounced in the acute phase of recovery (Hurford et al., 2013; Nys et al., 2005b). Moreover, if recovery occurs during this period, it is generally maintained in the years following the stroke (Nys et al., 2005b). Given this state of

affairs it is not surprising that researchers have begun to explore the possibility that natural recovery could be enhanced through training interventions. In the next section, the literature is reviewed concerning interventions aimed at facilitating natural recovery in the domains of current interest. The distinction between chronic and acute phase is maintained as the method for organising the studies given the substantial differences in prevalence and recovery rates during these periods.

There is a long history of attempts to improve cognitive performance via training interventions. In the context of working memory training, Klingberg (2010) makes a broad distinction between implicit and explicit training programs that form the basis of inclusion and exclusion criteria for selecting studies.

Training on motor and perceptual tasks in animals leads after hundreds of trials to enhanced performance, with concomitant changes in synaptic connectivity in both sensory and motor areas..... This type of perceptual and motor training might be called implicit because improvement is based only on repetition, feedback and often gradual adjustment of the difficulty. By contrast, teaching of strategies to improve WM tasks such as rehearsal, chunking and meta-cognitive strategies are explicit in that they are conscious strategies for handling the material. The question is whether implicit WM training might lead to durable neuronal changes in WM-related areas in the same way as perceptual training does for neurons of the visual cortex. (p. 317-318)

The distinction between implicit and explicit training complements the restorative – compensatory distinction outlined in Chapter 1. Implicit training focuses on restoring normal neural functioning whereas explicit training accepts that neural damage has occurred and training is aimed at producing and enhancing compensatory behaviours for the behavioural

deficits. In the following review, studies that deal only with implicit training are considered because of their presumed linkage to changes that are occurring in the brain.

One other crucial aspect of the training literature that needs to be addressed briefly (it will be addressed at length in the next chapter) is the issue of transfer of training. In the mainstream cognitive training studies one or more cognitive tasks are used for training. Repeated practice on these tasks over a specified time period constitutes the training regime. The expectation in such studies is that performance on the training task will improve with practice. The major interest in these studies, however, is the extent to which training transfers to other non-practiced tasks. The rationale here is that if a cognitive domain is enhanced through training, that enhancement should be reflected in tasks that tap that cognitive domain even if the specific tasks have not been practiced. Thus, a study that evaluates the outcomes of a training regime will usually collect pre-test post-test data on tasks that are assumed to measure the same domain as the training task (transfer tasks). The assumption is that there will be enhanced performance on the post-test results. The studies also include a set of cognitive tasks (control tasks) that measure one or more of the non-practiced domains. Since training has not involved these domains, the expectation is that no pre-test post-test differences should emerge. If there is improvement due to increased familiarity with the tests, then improvements will be greater in the transfer tasks than the control tasks.

However, it must be kept in mind that there is one essential difference between training studies aimed at enhancing abilities in cognitively unimpaired individuals and those who have suffered a stroke. In the case of the former, the cognitive system is intact and training is aimed at improving and enhancing well-established processes. In the case of a stroke, the cognitive system has been damaged and training is aimed at re-establishing pre-existing functioning cognitive processes. That is, training in the case of patients with stroke is about re-learning, rather than learning (Kleim, 2011).

### **2.6.1 Cognitive Training in the Chronic Phase following Stroke**

Training programs in cognitive rehabilitation have been the subject of a number of reviews. Cicerone et al. (2005, 2011) for example, conducted two reviews covering studies from 1998 to 2008. These studies explored training in executive functions, attention, memory, language, and perception, and included studies dealing with stroke and with traumatic brain injury. In the 2005 review search processes produced 118 studies of which 87 met their inclusion criteria. In the 2011 review, of the initial pool of 141 studies that had been published between 2003 and 2008, 112 met inclusion criteria. They concluded that with traumatic brain injury as the relevant population, there was evidence supporting the effectiveness of training in executive function, attention and memory. In the case of stroke, training was deemed to be effective only in visual rehabilitation for those who had suffered a right hemisphere stroke and for language rehabilitation for those suffering a left hemispheric stroke.

The Cicerone et al. (2011) review represents the latest instance where multiple cognitive domains have been reviewed. Recent reviews have been more focused on particular domains. Two recent Cochrane reviews have examined attention training in patients with stroke (Loetscher & Lincoln, 2013), and training on executive function outcomes in people with stroke and other forms of non-progressive acquired brain damage (Chung, Pollock Campbell, Durwood, & Hagen, 2013). As is the standard procedure in Cochrane reviews, only randomised control trials were considered for inclusion.

In the case of attention training, Loetscher and Lincoln (2013) identified 2785 studies using their search criteria of which 2721 were immediately rejected. Of the 64 studies 58 were also rejected for a number of reasons such as the incorrect population, training did not focus on attention, the study did not involve randomised control trial. In the remaining six studies that met their inclusion criteria, attention training was compared to care-as-usual

controls in the areas of alertness, selective attention, divided attention, sustained attention, and spatial attention with a global measure of attention also being calculated. A meta-analysis of these studies found that there was no statistical support for the long-term effect of training on global measures of attention, nor in the specific domains of alertness, selective attention, sustained attention or spatial attention. However, there were statistically significant beneficial effects of training in the divided attention tasks. In short, patients with stroke improved in their ability to divide their attention between two or more tasks.

Training effects on executive functioning in stroke, acquired brain injury and other non-progressive brain injuries were examined in a review by Chung et al. (2013). Their initial searches produced 8,280 studies of which all but 121 were readily excluded. Of the 121 potential studies, the application of their inclusion criteria resulted in 13 studies that met the requirements for a meta-analysis. The aspects of executive function that were assessed were concept formation, planning, flexibility, and inhibition. They also included working memory as another dimension of interest. Based on their meta-analysis, they concluded that training had little effect on recovery of any components of executive functioning and working memory. However, it should be noted that only four of the studies involved purely patients with stroke and only one study (Westerberg et al., 2007) used training or outcome measures that dealt with one of the domains being explored in the current thesis, namely working memory. This study will be described later.

Poulin, Korner-Bitensky, Dawson and Bherer (2012) also reviewed interventions aimed at improving executive functions but used less stringent inclusion criteria than the Cochrane reviews. Their initial search criterion limited the population to patients with stroke resulting in 1,535 potential studies of which all but 26 were readily excluded. Their review found 10 studies that met their inclusion criteria. They concluded that there was little evidence for the effectiveness of training. It should be stressed, however, that seven of these

studies involved training on compensatory tasks (paging system, goal setting and problem solving strategies). Only three of the studies involved cognitive training aimed at restoration of damaged domains. The three relevant studies that involved cognitive training are reviewed below. A fourth study not reviewed by Poulin et al. (2012) is also included.

Stablum, Umilita, Mogentale, Carlan and Guerrini (2000) studied the effects of a training regime on nine participants who had suffered a haemorrhagic stroke. Initial neuropsychological assessment was conducted 7 months after the stroke. The stroke group's scores on several executive functioning tests were significantly below those of a no-contact age- and gender-matched control group. The training involved 5 hours (1 hour per week for 5 weeks) of dual task training which was aimed at improving working memory/ executive functioning domains. The primary task involved the brief presentation of a pair of letters to the left or right of a fixation point on a computer screen. The task was to indicate which side of the screen the letters appeared on, and participants were to do this as quickly as possible by pressing a key on the computer keyboard. In the dual task condition the non-speeded secondary task was to simply say whether or not the two letters were the same. There was a consistent linear improvement ( $R^2 = .95$ ) in dual task performance across the five training sessions and clear improvement between pre-test and post-test scores for the stroke participants. There were no pre-test post-test differences for the control group. The improvement in dual task performance was maintained at 12-month follow-up. There were also pre-test post-test differences in some of the non-trained cognitive tests such as the PASAT, Continuous Performance Task, Trail Making Test, and backward Digit Span, but no differences on other tests. The authors interpreted these latter findings as being indicative of transfer. In short, dual task practice led to improvements in performance on the training task across sessions, and there was evidence of transfer to other executive tasks.

Vallat, Azouvi, Hardisson, Meffert, Tessier and Pradat-Diehl (2005) devised a theory-based intervention in a single case, multiple baseline study for a person who had suffered a stroke. Initial neuropsychological testing had indicated that all cognitive abilities were intact except for working memory. Consequently, an intervention based on Baddeley's (1986) model of working memory was created, and training began 14 months post-stroke. The training tasks were designed around the processes associated with the phonological loop and central executive components of the working memory system (Baddeley, 1986), but did not involve the tasks that have typically been used to assess the two components, for example, forward digit span (phonological loop) and backward digit (central executive). The training regime involved three 1-hour sessions each week for 6 months. Three different types of outcome measures were employed: those that were directly related to the model (forward and backward digit span); other non-trained complex working memory tasks (transfer); and other non-trained cognitive tasks that did not involve working memory (control tasks). The patient's performance at baseline and post-test was compared to a control group of 10 people who were matched on demographic characteristics. Changes occurring during the training phase were not reported; only the pre-test post-test differences on the outcome measures were reported. The training was effective in that there were statistically significant improvements on the targeted phonological loop and central executive tasks. Significant pre-test post-test changes on the complex span tasks that are underpinned working memory (Unsworth & Engle, 2007a), but no such changes on the control tasks, indicated that transfer had occurred.

Westerberg et al. (2007) investigated working memory training in patients with stroke living in the community who had suffered their strokes 20 months, on average, prior to testing. Of the 24 people originally contacted, three declined to participate and three did not complete follow-up testing, leaving 18 participants who completed all phases of the experiment. The inclusion criterion for the patients with stroke was a self-report statement

that they had problems in attention. No neuropsychological test data were reported to confirm the presence of an attentional impairment, so it is not clear that any of these participants were in the clinical range in the working memory/attention domain prior to the commencement of training.

The 18 participants were randomly assigned to either a working memory training group or a no-contact control group. The training regime itself was computer controlled with the participants doing the training in their own homes. The training regime involved a number of different training tasks that had four common characteristics: the maintenance of multiple stimuli at the same time; short delays in which the stimuli had to be maintained in working memory; the maintenance of serial order of the stimuli; and increases in difficulty as training progresses. The training regime consisted of 40 minute sessions, five times a week for 5 weeks. In reporting the outcomes of the study there was no description of learning during the training phase so it is not possible to determine whether or not performance on the training task improved.

The effectiveness of the training involved pre-test post-test measures on a neuropsychological test battery that consisted of direct measures of working memory, other tasks assumed to rely on working memory (transfer), and control tasks where working memory was not involved. The training group demonstrated improvement on both the direct working memory tests and the transfer tasks. There was no improvement on the non-working memory tests. This study represents one of the strongest sources of evidence for training effects facilitating the recovery process.

In the most recent study exploring working memory training in patients with stroke, Nordvik, Schanke, Walhovd, Fjell, Grydeland and Landro (2012) explored the relationship between working memory training and white matter changes in a single-case study of a 60-year-old man. Four and a half months after his stroke the participant underwent his first

cognitive evaluation. The evaluation indicated that he had an impairment in working memory (5<sup>th</sup> percentile), but other cognitive domains such as visual and verbal memory, executive function and speed of processing were largely unaffected. Subsequently, he undertook a five-phase training program (ABABC design) that consisted of a baseline measure; a computerised training program that addressed visuo-spatial, attention, memory, executive functions and problem solving; a subsequent period of no training; then two periods of further training where the first program was run again; and a final phase where Cogmed was used to specifically target the working memory system.

The outcome measures were a series of neuropsychological tests that measured working memory, immediate and delayed memory, executive skills, visuo-spatial processing, abstract reasoning and processing speed. These tests were administered six times, at the beginning of each training phase and at the conclusion of training 22 months post-stroke. Figure 2.3 shows that performance on all the outcome measures increased in a roughly linear fashion across the six testing sessions for all domains except for working memory. With working memory, the first training session appeared to produce little change. However, in the last two phases, particularly when Cogmed was introduced, there were substantial improvements in performance such that by the end of training his working memory scores were at the 50% percentile.

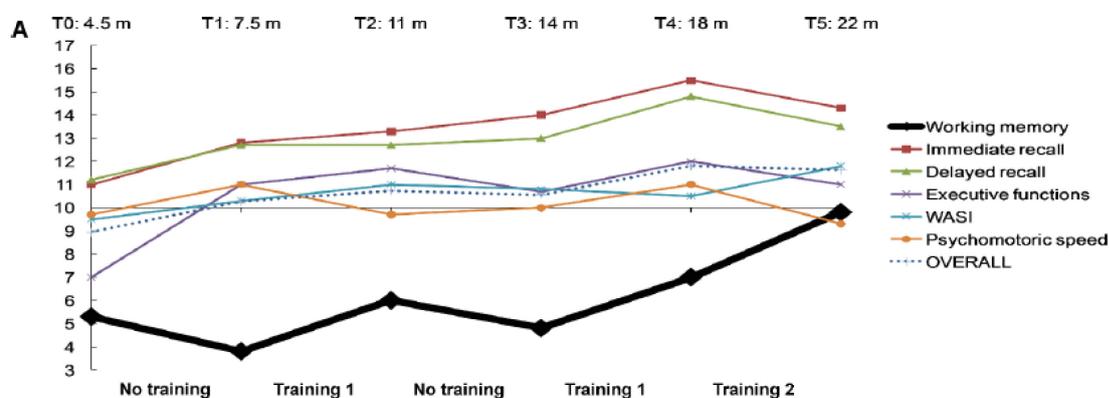


Figure 2.3. Training outcomes in Nordvik et al. (2012) study.

The results in Figure 2.3 are interesting because they highlight some of the issues associated with training studies involving stroke. For a start, the patient has an impairment in one domain but not in other domains. The initial training program involved training across multiple cognitive domains that were represented in the neuropsychological test battery. In the intact domains there was clear improvement across the six testing sessions on the neuropsychological tests, which reflect standard practice effects. Moreover, in the early phase of training there was very little improvement on the working memory tests. This could be interpreted as the participant not responding to an effective program. Alternatively, it could be that the working memory training program itself was ineffective regardless of an initial impairment or not. These alternative explanations could be eliminated if normative data on learning rates on the various tasks were available. In other words, if a control group had participated in the study doing only the neuropsychological tests and not the intervention, practice effects on the neuropsychological tests would be known. With this information, less ambiguous inferences could be made about the presence or absence of training effects exhibited by the stroke patient.

### **2.6.2 Section Summary**

There are a number of conclusions that can be drawn from the studies reported here. Firstly, there are five large review studies (Chung et al., 2013; Cicerone et al., 2005, 2011; Loetscher & Lincoln, 2013; Poulin et al., 2012) dealing with cognitive training effects that have used wide search terms to generate potential studies for consideration. Each of those studies concluded that there are actually very few studies that assess the effectiveness of implicit training regimes in the areas of concern in this thesis.

The few studies that do exist (Nordvik et al., 2012; Stablum et al., 2000; Vallat et al., 2005; Westerberg et al., 2007) are promising in that all studies show learning and transfer

effects that are specific to the domains targeted by the training. The implication of these outcomes is that cognitive training produces positive outcomes.

There are some issues however. Given the prevalence and recovery data described earlier in the chapter, only 30% to 40% of participants in the chronic phase of recovery have cognitive impairments. If cognitive training is to be useful in the clinical setting it must work with people who are actually impaired. In two of the reported studies the data that would indicate whether or not the participants were impaired were not presented. A second assumption being made is that training on a task will lead to improved performance on that task over time. It is only if this occurs that it makes sense to look for transfer effects. If there is no learning going on during training then there is no expectation that there will be any transfer effects. This means that where pre-test post-test differences on transfer tasks do not emerge, in the absence of training information, one cannot be sure whether there is no transfer or if there is no effect of training. Again, two of the studies do not report training data.

The case study described by Nordvik et al. (2012) illustrates the need to establish impairment in the first instance, to report training outcomes, and to have normative data with which to compare performance changes. Their stroke patient had no impairments on a number of domains and repeated administration of the same battery of neuropsychological tests that targeted multiple domains showed improvement in the unimpaired domains, presumably reflecting practice effects. Since there was no initial impairment, and therefore no need for recovery, the improvement that their participant exhibited could be interpreted as what might be expected in a normative sample had it been employed. However, to be unequivocally certain that these effects are practice effects, a control group would need to be exposed to the same testing regime without the intervention, and show the same degree of improvement.

The true test of training effectiveness is only possible where an impairment has been identified and normative data is needed to make such a comparison. In the Nordvik et al. study, the presentation of the test data for all five sessions showed that the early interventions produced no such improvement in the impaired domain. In other words, while there was substantial learning going on in the unimpaired domains, there was little learning going on in the impaired domains. In the absence of normative data one cannot decide if the lack of learning was abnormal or not.

### **2.6.3 Cognitive Training in the Acute Phase following Stroke**

The review above involved training studies where training occurred in the chronic phase of stroke. Three of the above reviews considered cognitive training during the acute phase of stroke. Cicerone (2011) stated that no studies involving training in the acute phase could be found in the 112 studies reviewed. Chung et al. (2012) made a similar statement that they could find no training studies that commenced in the acute phase. Poulin et al. (2012) also specifically searched for training studies in the acute phase and could find none. The conclusion to be drawn here, is that there are no published studies where direct cognitive training of the functions and domains typically assessed via neuropsychological tests has commenced during the first weeks after stroke. This represents a second gap in the literature that the current experiments are designed to address.

## **2.7 Recovery versus Training in Chronic Phase**

The literature shows contrasting recovery and training effects during the chronic phase. In the case of recovery, the repeated administration of neuropsychological tests shows that there is only modest improvement in performance across time. In contrast, the training studies show evidence of substantial improvements in performance on both the trained tasks and on transfer tasks. While there are differences in outcomes, the basic procedures are similar in that a set of measures are repeatedly sampled in each case, which begs the question

as to why repetition of one set only produces modest learning and the repetition of another set shows strong learning. One obvious difference is in the number of repetitions. For example, in the recovery literature the neuropsychological tests are administered on two or three occasions at most, with these repetitions being widely spaced. In the case of the training tasks, the tasks are repeated many times for an extended period of time. Consequently, the slow behaviour change associated with recovery and the robust improvements associated with interventions can plausibly be attributed to the degree of exposure and engagement with the relevant tasks.

While the number of repetitions may be an important factor, there are other differences between the two situations. The tasks used in the recovery studies and in the training studies are quite different and serve very different functions. The tasks used in the recovery studies are primarily those used in neuropsychological assessment. These tasks have well-established protocols based on the performance of large samples who are representative of the general population. The psychometric validity and reliability of these measures have also been subject to much scrutiny. However, these tests are designed primarily for one-off administrations that assess current intellectual functioning. Most of these tests are not designed for repeated use. In fact, changes in performance in re-administration are seen as a nuisance with the optimal and preferred outcome being that repeated administrations produce no improvement in performance (Green, Strobach, & Schubert, 2014). However, classic test theory does acknowledge the reality of practice effects and, as will be described in some detail later, does have mechanisms that allow for true practice effects to be discriminated from normal variability in test performance (Strauss, Sherman, & Spreen, 2006).

In the case of many of the training studies, the training tasks are based on, and tightly linked to specific cognitive theories. Thus, the training tasks in the Stablum et al. (2000), study involved a dual task methodology that is widely used in mainstream exploration of

attention and working memory (Baddeley & Hitch, 1974). In the Vallat et al. (2005) study, the training tasks were based on the processes thought to underpin the phonological loop and central executive components of Baddeley and Hitch's working memory model. The Westerberg et al. (2007) and Nordvik et al. (2012) studies employed working memory tests from the Cogmed training program. These tests are visual and verbal variants of simple span tasks where items are briefly presented and have to be recalled in a specified order (forward, backward, or random cued). These are standard tasks that are widely used to measure short-term and working memory capacity (Unsworth & Engle, 2007).

While these cognitive tests have been widely used, the psychometric properties of the tests are not of prime importance in the cognition literature and are often not reported. Moreover, because these tasks are not used for assessment purposes, there are no norms for what constitutes good or poor performance on these tasks. In short, these tasks do not have the "history" of the neuropsychological tasks when it comes to evaluating cognitive abilities.

Training tasks also contrast with neuropsychological tests because it is expected that repetition effects should lead to enhanced performance. That is, sustained engagement with these tasks is enacted specifically to produce changes in performance. The absence of such changes would be considered aberrant and would be grounds for abandoning the use of the task. In the literature cited, the cognitive tasks do produce effective improvement in performance with repeated exposure.

One other possible explanation for the differential rate of behaviour change in the recovery and training studies may be due to the potency of different types of tasks in producing behaviour change. It might be the case that repetition effects in the neuropsychological tests are very weak, and the tests may have been selected in part for this property. Repetition in the cognition literature, on the other hand, is seen as a fundamental means of learning, and it may be the case that the cognitive tasks used in the training studies

are those that are most sensitive to repetition effects. In any consideration of training studies one needs to consider the potency of the training tasks.

One final issue between recovery and training concerns the ability to distinguish recovery effects from training effects when the only manipulation involved is the repetition of a set of measures. While in the training tasks repetition is seen as a basic means to induce learning, the repeated administration of neuropsychological tests could also be considered as a form of intervention, producing the same improvement in performance across sessions as has been observed in the training studies. Consequently, there is no straightforward way of discriminating recovery from training. Training and recovery are more easily distinguished when performance on non-trained transfer tasks is examined in the training studies, but with simple repetition of tasks the distinction between recovery and training is blurred.

### **2.8 Characteristics of Effective Training Programs**

The research around cognitive training effectiveness has been conducted in the context of attempting to improve working memory capacity in both normal participants and those populations that are associated with a perceived cognitive deficit. Over the last decade, perceptions of training effectiveness have evolved from initial confidence and enthusiasm to high levels of scepticism (Shipstead, Redick, & Engle, 2012). The debate centres around three issues: concept measurement, generalisation, and appropriate controls, but it also addresses the issues involved in the delivery of such training programs.

The initial promise of working memory training was based on a number of studies that suggested that training on working memory tasks led to increased performance in abstract reasoning (Klingberg, Forssberg, & Westerberg, 2002), attention (Chein & Morrison, 2010; Klingberg et al., 2002), and in reducing ADHD symptoms (Klingberg et al., 2005; Mezzacappa & Buckner, 2010).

As Shipstead et al. (2012) argue, the basic assumption being made in this literature is that training improves working memory capacity, and that subsequent improvements in other cognitive domains are due to the fact that these other tasks utilise working memory capacity. There is no doubt that training on a working memory task leads to improved performance on that task, if the task remains difficult throughout training. The issue is the extent to which this training measures working memory capacity rather than other task characteristics, and to what extent such training generalises to untrained tasks. If training on a task improves performance on that task, then one would expect that other measures of working memory should show an improvement even if they have not been trained. This is known as near transfer. Beneficial training on working memory tasks should extend further than just other working memory tasks, if working memory is related to a second cognitive construct. This is known as far transfer. The initial enthusiasm for working memory described above was based on such demonstrations of near and far transfer.

Some disenchantment with working memory training has begun to emerge for two reasons. Firstly, there are an increasing number of studies where training effects are not found; and secondly, issues have been raised about the theoretical and methodological underpinnings of the training studies.

Failures to replicate near and far transfer effects are problematic. While a number of studies show large transfer effects (Klingberg et al., 2002), there are an increasing number of studies where no transfer effects have been found (Gathercole, Dunning, & Holmes, 2012). Melby-Lervåg and Hulme (2012) conducted a meta-analysis of 30 working memory training studies where performance on the training group could be compared to that of a control group. The results of this meta-analysis indicated that near transfer effects were real, but they were not maintained for any great length of time. However, there was no evidence supporting

far transfer to other cognitive tasks like nonverbal and verbal ability, inhibitory processes in attention, word decoding or arithmetic.

The discrepancies in outcomes that motivated Melby-Lervåg and Hulme's (2012) meta-analysis have also prompted a task-analysis approach to understand the training effects from both theoretical and methodological perspectives. Shipstead et al. (2012) have argued that the training literature suffers because it does not have an empirically based account of working memory, it fails to adequately explore possible confounds that might account for training effects, and does not address the mechanisms of training.

Regarding the first issue, they make the distinction between the training task and theoretical construct, which is deemed to underpin task performance. The issue here is that performance on a task can improve for a number of reasons because all cognitive tests are multi-faceted. Given training on a complex span test where improvement is observed, it is possible that working memory capacity can be enhanced. Such improvement, however, could simply be the result of greater familiarity with the test, repetition effects or adopting a more effective strategy for doing the task.

In the experimental literature where the relationship between working memory capacity and intelligence has been widely studied, the constructs of working memory capacity and intelligence are always tested using multiple tests. The assumption here is that each test will have some component that reflects working memory capacity and some components that are unique to the task. By using multiple tests, one can isolate the latent common component in each domain (working memory latent component and intelligence latent component) and look at the relationships between the latent constructs. If one uses a single working memory task for training and transfer is observed, one cannot be certain if it is the working memory component that is causing the transfer, or whether the two tasks share some other properties in common. Shipstead et al. (2012) note that the majority of working

memory training studies train a single task only. They argue that transfer in many instances may reflect task-specific practice, not changes in the underlying construct. Few training studies exist where training has used multiple tasks to derive changes in latent variables and consequently it is virtually impossible to tease apart domain improvement from improvement due to other task characteristics.

This observation about the various task components impacting on the interpretation of training effects is even more problematic in the case of stroke. Given the expectation that natural recovery leads to improved performance, a somewhat disappointing conclusion from the Shipstead et al. (2012) arguments is that it may well be near impossible to separate the effects of training from natural recovery.

Methodological problems also arise as many studies do not use a control group, or if they do, it is typically a passive, no-contact control group. In such control groups the participants do not participate in the experiment except for taking pre-test and post-test measures. Improvement in the training group under such circumstances may well be due to training effects, but placebo and changes in behaviour that occur by simply being observed (the Hawthorne effect) could also produce changes in performance. Some studies have controlled for these effects by adding an active control group who go through the same training schedule but with subtle differences in the training regime. For instance, the control group may go through the training process, but the level of task difficulty remains at the easy level of the spectrum. There are few studies that employ an active control group.

As a result of a number of working memory training reviews it seems that an adequate test for a training program's effectiveness depends on demonstrating both near and far transfer effects. To do so adequately involves training using multiple tests of a domain, the use of appropriate control groups, adaptive training procedures to ensure increasing levels of task difficulty, using sufficient numbers of participants to ensure statistical power, and

random assignment of participants to conditions. While the above reviews have been done in the specific realm of working memory training, the comments would apply to training in any cognitive domain. Needless to say, none of the training programs involving patients with stroke reviewed earlier meet this standard, so evaluating their effectiveness is open to alternative interpretations.

Based on the training literature, the weight of opinion is that improvement in working memory capacity as a result of working memory training regimes has yet to be demonstrated convincingly. However, as Gathercole et al. (2012) note, conducting gold standard research is “...time consuming and expensive and until the existing evidence indicates that the intervention passes less stringent empirical tests, the intervention is risky.” (p. 201) Gathercole et al. (2012) argue for a pragmatic approach that is aligned with US National Institute of Health guidelines on conducting clinical trials (<http://clinicaltrials.gov/ct2/info/understand/>). One starts with proof of concept, then studies designed to refine and calibrate the intervention while controlling for possible confounds. It is only after these studies have been conducted that full-scale randomised control trials are initiated.

Gathercole et al. (2012) make one further point that is relevant to current concerns. They concede that while less than gold class methodologies are problematic if training effects are found, the absence of any training effect can be just as informative, if not more so, than its presence. From a practitioner perspective it is probably more important to know when training is not effective than what precisely caused the improvement when training is effective.

## **2.9 Literature Review Conclusions and Aims of Study**

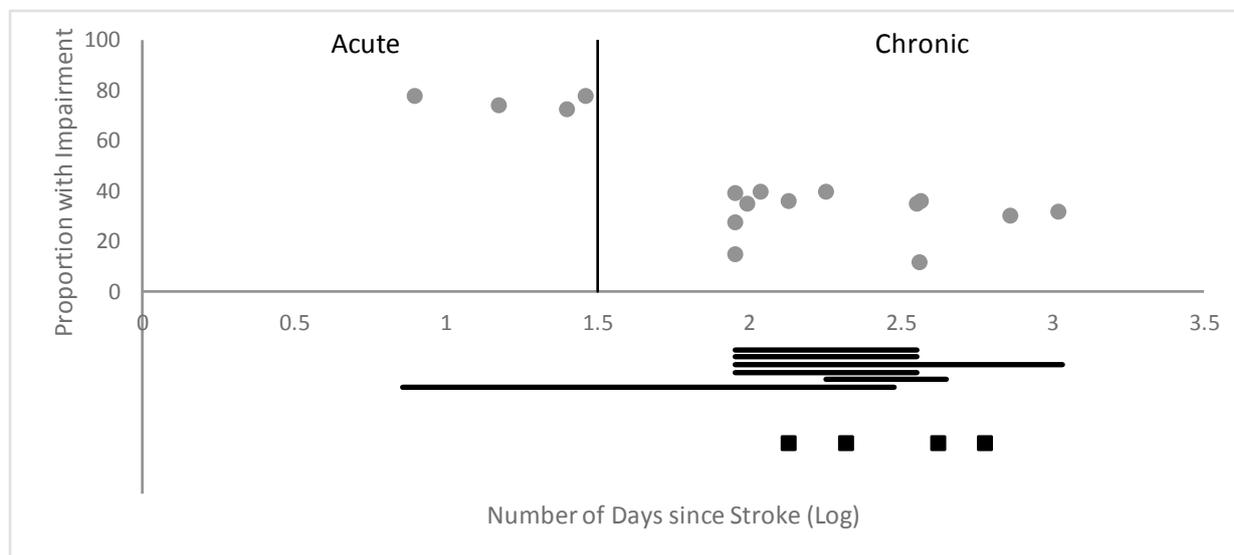
### **2.9.1 Conclusions**

The chapter began by specifying that the focus of this thesis would revolve predominantly around the cognitive domains of executive functioning, speed of processing, attention, working memory, and visuo-spatial skills because these are the cognitive domains that are most frequently affected by stroke (Cumming et al., 2013). Impairments in these domains are also not obvious and such impairments are often overlooked (Planton et al., 2012). The failure to detect such deficits has clinical ramifications because the degree of impairment in these areas is predictive of long-term outcomes in functional areas of behaviour (Tatemichi et al., 1994).

In the clinical context, cognitive functioning is usually evaluated in the early phases post-stroke by the use of cognitive screening tools such as the MMSE and the MoCA. The utility of these tools is based on a small number of items that can be administered quickly and easily while the patient is ill and has limited powers of concentration. The downside of such tools is that they can identify the presence of a cognitive impairment but they cannot reliably determine which aspect of cognitive performance is impaired (Nys et al., 2005a). The second phase of evaluation, if undertaken, is usually done months after the stroke when neuropsychological test batteries that cover most aspects of cognition are conducted. Such neuropsychological tests are more sensitive than the screening tools in detecting cognitive impairment, and in determining which cognitive domains survive a stroke and which are adversely affected.

Figure 2.4 presents the essential outcomes of the key features of the literature review. The x-axis represents the number of days since the stroke as measured in log units and represents the period between 1 day and 3 years post-stroke. The area above the x-axis presents the estimates of the number of patients with stroke in each study who show an impairment in at least one domain. The horizontal bars below the x-axis depict the time course of the studies that have explored recovery, and the squares reflect the time when the

initial assessment of cognitive function was undertaken in the training studies. The vertical line divides acute and chronic phases.



*Figure 2.4.* Summary of prevalence, time-course of recovery, and initiation of training outcomes of the literature review

There are a number of points to be made regarding the prevalence data in Figure 2.4. Firstly, there are many more studies where cognitive assessment was undertaken in the chronic phase than in the acute phase. In fact, there are very few studies that have explored cognitive performance in the first few weeks after stroke. However, the literature consistently shows that impairment is more prevalent in the acute phase of stroke (70% - 80%) than in the chronic phase (20% - 40%). In the acute phase more people show a cognitive impairment in at least one cognitive domain and more people show multiple impairments than is the case in the chronic phase. This pattern implies that people can and do recover from their stroke. Which particular domain is most affected (not depicted in Figure 2.4) differs from study to study, due in part to definitional issues. Domains such as executive functioning, speed of processing and working memory impairments, however, are regularly observed, and all

cognitive domains can be affected. In the case where performance of groups of patients with stroke have been compared to control participants (Jaillard et al., 2009; Tatemichi et al., 1994), patients with stroke show significantly lower performance across most tasks compared to control participants.

Studies that have explicitly examined the cognitive recovery process have confirmed that recovery does occur and it occurs over time. Recovery seems readily observable when initial testing is done in the acute phase and follow up is done in the chronic phase (Nys et al., 2005b, but when all testing is conducted in the chronic phase, behaviour is more stable and changes are harder to detect. One gap in the literature that is evident in Figure 2.4 is that searches of the published literature have not produced studies where multiple neuropsychological testing has been conducted during the acute phase. Logically there must be changes during this period, but the author has not been able to find any studies that track changes over this period. Being able to track such progress, or the lack of it, is important to the clinician as the data appears to suggest that changes during this period are highly predictive of long-term outcomes (Tatemichi et al., 1994).

There are a number of studies that have addressed the issue of whether or not the natural recovery process can be enhanced by implicit training interventions (Nordvik et al., 2012; Stablum et al., 2000; Vallat et al., 2005, Westerberg et al., 2007). In all these studies, however, diagnosis has been made and training begins in the chronic phase. These studies also differ from the previous prevalence and recovery studies in that they involve small numbers of patients with stroke, often a single patient. The outcomes are encouraging because there is evidence that training does lead to enhanced performance and that transfer of learning can be observed (Stablum et al., 2000; Vallat et al., 2005; Westerberg et al., 2007). However, those few studies that have addressed post-stroke training during the chronic phase of recovery sometimes do not indicate whether or not there are impairments in the domain

that is being trained (Westerberg et al., 2007), and sometimes do not report on behavioural changes on the training task throughout the training period (Vallat et al., 2005; Westerberg, 2007). Without these two pieces of information it is difficult to determine when any improvement in performance reflects enhancement of an impaired cognitive function or standard learning effects in an undamaged cognitive function. In the instance where there is a lack of transfer, it is quite possible that training was unsuccessful.

The second major outcome of the literature review as depicted in Figure 2.4 is that literature searches have failed to produce training studies where direct cognitive training in the areas of concern to this thesis has begun during the acute phase post-stroke. This represents another large gap in the literature. Based on the evidence presented in both Chapter 1 and Chapter 2, the acute phase is where improvement is most readily detected and interventions are most likely to have maximal long-term beneficial effects. Therefore it is imperative to begin exploring cognitive training during the early weeks following stroke.

### **2.9.2 Aims and Objectives.**

The literature review suggests that there are two large gaps in the effects of stroke on cognitive functioning in the specified areas of cognitive performance. Firstly, there is no literature that tracks the changes in cognitive performance during the first 4 to 6 weeks following stroke. The second is that there are no training studies that have been initiated during this same period. Given that the research shows that cognitive improvement during this period is predictive of long-term outcomes, the first general aim of the thesis is to provide empirical data that begins to address both gaps in existing knowledge. The intent is to explore changes in behaviour through the repeated administration of a battery of neuropsychological and cognitive tests while patients with stroke are in hospital in the early stages of recovery from their first stroke.

With regards to the recovery process, the intent is to track this via repeatedly administering a number of neuropsychological tests. As indicated earlier, neuropsychological tests were not designed for repeated use, and certainly not for multiple repetitions over a short-time frame. If these tests are subject to practice effects, those effects should be most apparent under these conditions (Calamia, Markon, & Tranel, 2012; Darby, Maruff, Collie, & McStephen, 2002) with the consequence that distinguishing between practice effects and natural recovery could be problematic.

It is not possible to conduct a training study that meets the required conceptual and methodological standards as outlined in the critiques of working memory training studies (Melby-Lervåg & Hulme, 2012; Shipstead et al., 2012). Given the projected time constraints, it is not possible to use an appropriate baseline – intervention – posttest training design that involves near and far transfer tasks.

What is possible to achieve under the projected constraints is to investigate the effects of repeating a battery of tests multiple times across a short timeframe to explore changes in performance. By combining both neuropsychological and cognitive tests in the battery, it should be possible to verify the presence of deficits in the aftermath of a stroke, and to observe any improvements in either impaired or unimpaired tasks. The inclusion of neuropsychological tests capitalises on the fact that such tests have well-established assessment credentials, but little is known about the effects of administering such tests multiple times over a brief period. Including of a set of cognitive tests that are theoretically driven, helps examine the potency of different types of tasks in producing improvements in performance, which might form the basis of future interventions. Thus, following the suggestion of Gathercole et al.'s (2012), the inclusion of these tests might form the proof of concept evidence that underpins effective intervention regimes. The down-side of using these cognitive tasks is that they do not have the psychometric properties of the neuropsychological

tests, or the well-established norms with which the performance of patients with stroke can be compared.

Thus, the objective of the current study is to explore changes in behaviour that are associated with multiple administrations of a battery of neuropsychological and cognitive tasks under the same conditions. As a first step, the cognitive tests need to be evaluated along the same dimensions as the neuropsychological tests. That is, the psychometric properties of the cognitive tests need to be established. Because there is limited existing data relating to administration of either neuropsychological or cognitive tasks multiple times over a short period, the second phase of the research examines the degree of learning associated with each task. Moreover, these data need to be collected from an appropriate sample so that they can serve as preliminary normative data by which patients with stroke' performance can be compared. Having established a psychometrically sound battery, the third and final stage explores performance changes of those who have been admitted to a rehabilitation ward after suffering a recent cerebrovascular insult in order to assess the utility of the developed battery. Utility would be defined by four outcomes: a) the ability to demonstrate stable behaviour, b) the ability to be able to differentiate impaired from unimpaired performance, c) the ability to detect regular and consistent improvements across test sessions, and d) for differences between neuropsychological and cognitive tests to be present.

Three different general approaches could be adopted to explore the utility of the battery in assessing behaviour change. The first is a group approach where performance of a group of patients with stroke is compared to a group of matched healthy people. The group approach is not being used in the current study for two reasons. Firstly, the patients with stroke are likely to have multiple impairments but also some preserved abilities. The profile of which domains are affected and which are not will differ from person to person. When learning is evaluated on a particular measure, it is likely that any large group of patients with

stroke will involve some who have an impairment and some who do not have an impairment. If those who do not have an impairment show normal learning on a task, and those that have an impairment do not, it is still the case that on average the stroke group might show significant learning, although not as substantial as in a control group. Matching people on profiles would mean an inordinate number of participants would be required.

The problem of matching patients with stroke is also exacerbated by the fact that stroke is frequently accompanied by comorbid conditions such as Parkinson's disease, dementia, depression, diabetes, hypertension, and other illnesses, all of which can have a detrimental impact on cognitive performance (Bohnen et al., 2014; Bott et al., 2014; Caughey et al., 2010; Gallacher et al., 2014). This means that any group of patients with stroke is likely to be elderly, contain patients with different impairment profiles due to the stroke, and likely to have different comorbidity profiles as well. Controlling for all these factors is functionally not possible (Stuart & Rubin, 2007) and may not even be desirable (Frangakis & Rubin, 2002).

The second approach involves a case study methodology. The benefits of the case study approach in making inferences about the structure of cognitive systems based on the analysis of impaired performance has been strongly argued (Caramazza, 1986). Moreover, the intent of many neuropsychological case studies is to examine and elucidate some aspect of brain-behaviour relationships where the subject of the case study has suffered some form of neurological trauma. As such, detailed knowledge of the neurological and behavioural effects of the trauma are an essential requirement when interpreting the effects of the trauma. While a number of the tasks used in the current battery have been used to explore brain behaviour relationships in patients with stroke, the current focus of the research does not deal with such relationships. Before such relationships can be explored, it needs to be established that the tasks themselves are sensitive to change and can detect recovery. The current thesis,

with its emphasis on detecting change, is a necessary precursor to studies that explore factors that determine which domains are likely to be impaired and how recovery might be influenced by specific demographic and neurological factors like the location of the infarct.

Given the aims of this study, the case studies are more concerned with assessing and describing changes in behaviour than with explaining why the presence, absence or strength of behaviour changes were observed. The cases illustrate the different ways in which behaviour can change using the test battery. This means the cases are presented to show that the battery can produce stable data, can be used to differentiate impaired from non-impaired behaviour, can show different forms of behaviour change, and can demonstrate the potency of different types of tests. The use of these case studies is to argue that the current battery is a useful tool that the practitioner can use to evaluate behaviour change in the early days and weeks after a person has experienced cerebrovascular event.

In sum, the current research is primarily concerned with the process of evaluating behaviour change in the aftermath of stroke. The distinction between neuropsychological and cognitive tasks is maintained in order to address possible differences in the potency of different classes of tasks, and as a precursor to the development of intervention tasks that can be used to facilitate recovery in the acute phase of stroke.

## **Chapter 3: Development of Neuropsychological and Cognitive Tasks**

### **3.1 Chapter Overview**

Chapter 1 described the neurological repair, recruitment and relearning processes that came into play in the early weeks following strokes that affects the motor system (Wieloch & Nikolich, 2006). Plasticity changes were mirrored in equally substantial behavioural changes. Moreover, interventions that were initiated in the acute stage enhanced the neurological repair processes and produced lasting benefits in terms of behavioural recovery (Albert & Kesselring, 2012). Changes during the chronic phase were less substantial at the neurological and behavioural level (Duncan et al., 1992), and training that commenced during this phase was less effective (Cumming et al., 2011). Chapter 2 explored the same issues with respect to cognitive effects of stroke with the expectation that both behavioural and intervention effects would be similar to those observed in motor system recovery. The review of the literature showed that, where there was available data, the cognitive effects did indeed match those observed in the motor literature. There was clear evidence for the attenuation of cognitive impairments from the acute stage to the chronic stage (Nys et al., 2005b; Hurford et al., 2013), and that therapeutic interventions could have beneficial effects when initiated in the chronic stage (Stablum et al., 2000; Westerberg et al., 2007). There were two areas in which parallels between motor and cognitive recovery could not be addressed because there was no published data. Firstly, there were no available data that explored the recovery process during the acute phase of recovery. Secondly, there were no studies in which interventions had been initiated in the acute phase of recovery. These gaps in the experimental literature serve as the motivation for the following experiments and subsequent interventions.

In relation to recovery and intervention, a number of related issues have emerged. Firstly, the research indicated that during the acute phase patients were likely to have multiple cognitive impairments with some domains remaining relatively intact (Jaillard et al.,

2009; Nys et al., 2007; Tatemichi et al, 1994). For the intact domains, recovery is irrelevant and as such any behavioural change associated with repeated testing should be identical to such changes in a healthy population. In both instances, any improvement could be attributed to practice effects on the tests. In the case where an impairment was identified, one could then determine if there was subsequent improvement or not, and in the case where there was improvement could it be attributed to recovery or was it another instance of a practice effect? The ability to identify impairment was a crucial first step in interpreting subsequent changes in behaviour. Consequently, any research that explores recovery or intervention effects must have an assessment dimension to it (Ponsford, Sloan, & Snow, 2013; Stablum et al., 2000; Taylor & Broomfield, 2013; Wilson, 2009). The repeated administration of a battery of neuropsychological tests would provide information about initial levels of impairment and could also be used to track any subsequent improvement in the affected domains.

The second and third emergent issues involved possible interventions. The literature review in Chapter 2 indicated a lack of existing cognitive interventions that are aimed at enhancing restorative processes rather than compensatory strategies. While there are developed programs that explore specific domains like working memory (Cogmed), there are few programs that address multiple cognitive domains. Some of the “brain training” programs do attempt to address multiple domains, but the efficacy of such training tasks is questionable (Rabipour & Raz, 2012). Not only is there a lack of cognitive interventions, there are very few instances where an intervention has been underpinned by cognitive theory. The study by Vallat et al. (2005) is one exception to the rule, where the intervention was designed around Baddeley’s (1986) working memory model. Many authors in both mainstream cognitive training and cognitive rehabilitation literatures are currently asserting that all interventions should be supported by a strong theoretical rationale (Melby-Lervåg & Hulme, 2012; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, & Engle, 2010, 2012;

Wilson, 2009). This same assertion is frequently made in the cognitive rehabilitation literature as well (Ponsford et al, 2013; Taylor & Broomfield, 2013; Wilson, 2009)

The third emergent issue involves interpreting positive change in the case where impairment has been identified and an intervention has been initiated. The issue here is the ability to distinguish the effects of the intervention from those of natural recovery. To even begin to disentangle intervention effects from recovery effects, one needs to know what recovery looks like in terms of changes over time. As yet there is no empirical evidence on which to make any such judgement. Solving this problem is beyond the scope and resources of the current thesis. Instead, the more limited goal of exploring behaviour changes in impaired and intact cognitive domains using neuropsychological and cognitive tests to determine how much cognitive abilities improve in the early weeks after a stroke. A possible side benefit of this exploration is identifying which tasks are more or less suitable for use in the development of future intervention protocols.

In the following sections these issues are addressed with the aim of developing a series of tasks that can be administered in the acute phase of stroke. Issues in selecting and constructing the tests are described in the first section of the chapter, followed by a description of the actual tests themselves. Issues associated with the testing regime complete the chapter.

### **3.2 Neuropsychological Tests**

Both Wilson (2009) and Ponsford et al. (2013) assert that neuropsychological assessment is a primary starting point to any rehabilitation program. They describe assessment as a “*cognitive map*” where strengths and weaknesses can be identified, and more importantly, where impairments can be isolated. Not surprisingly, rehabilitation programs often use a neuropsychological framework, with impairments being categorised in domains

(executive functioning, memory, language, visuospatial) and modalities (visual, verbal, auditory).

Choosing a set of neuropsychological tests had to satisfy a number of constraints that were determined by practical, assessment and repetition considerations. The practical considerations are that the tests must be conducted in a hospital setting where testing conditions can be less than optimal. Secondly, one of the outcomes of stroke is that patients are often unable to maintain concentration for long periods of time, and they can become fatigued quite quickly (Jaillard et al, 2009; Nys et al., 2007). This means that test administration time should be as short as possible; tests that take a short period of time to administer are preferable to those that take a long period of time to administer. Thirdly, more reliable outcomes are obtained when there are additional data available. Tests that provide multiple measures of performance are preferred to those that produce a single measure. Finally, by minimising a task's motor components, cognitive assessment can still be extended to those whose stroke has impacted the motor system. Designing or selecting tasks that have minimal motor involvement would permit a wide usage of the battery.

The neuropsychological assessment literature reviewed in Chapter 2 suggests that a wide range of psychometrically sound tests should be administered to assess which particular cognitive domains are impaired. However, given the practical constraints detailed above, it was not feasible to take numerous measures of each cognitive domain. Therefore, the current battery was constructed so that the total administration time was less than 60 minutes, with priority being given to tasks that could be completed within 2 or 3 minutes, and that the tasks cover a range of cognitive domains.

The choice of domains tested was based on the literature review in Chapter 2. Consequently, executive functioning, working memory, continuous attention, language processing and visuo-spatial processing were selected as the domains of interest. Such

choices have the consequence that some domains were not tested at all. Examples of domains that were not tested were long-term memory, comprehension and vocabulary. The approach is to provide a wide, but shallow assessment of cognitive functioning. Consequently, the battery can be best considered as a screening tool for evaluating recovery rather than an in-depth assessment of how specific cognitive domains are affected by and recover from stroke. Such an exploration would involve more comprehensive batteries and extended testing sessions.

The research outlined in Chapter 2 was examined at the task level in order to identify tasks that met the constraints outlined above. This initial reading of the literature identified a number of tests that were frequently used across studies and were shown to be sensitive to stroke. The selected tests, their administration times and the purported cognitive domain that they assess are presented in Table 3.1.

Table 3.1

*Neuropsychological Tests, their Administration Time and Underpinning Cognitive Domains used in the Test Battery.*

Task	Administration Time	Cognitive Domain
Stroop Task		Executive Function (speed of processing, inhibition); Language
Read	1 Minute	
Colour Naming	1 Minute	
Colour-Word	1 Minute	
Digit Span		Working/Short-term memory
Forward Span	1-2 Minutes	
Backward Span	1-2 Minutes	
Rey Tangled Lines	2 Minutes	Visuo-spatial processing
Phonemic Fluency		Executive Function (Initiation); Language Processing
F	1.5 Minutes	
A	1.5 Minutes	
S	1.5 Minutes	
Semantic Fluency		Executive Function (Initiation); Language Processing
Animal Naming	1.5 Minutes	
Countries	1.5 Minutes	
Fruit & Vegetable	1.5 Minutes	

### 3.3 Developing the Cognitive Tasks

While the present thesis does not specifically explore intervention effects, the choice of cognitive tasks to be administered in the battery was influenced by the training and intervention literature. This literature suggests that decisions concerning the intended outcomes of the intervention are an essential first step. Such outcomes are seen to involve the restoration of lost function, encouraging anatomical reorganisation, using residual skills more efficiently, finding alternative means of attaining a final goal, modifying the environment, or an eclectic combination of these intentions (Wilson, 2009). The first two of these intentions are based primarily on neuroplasticity findings which imply that rehabilitation techniques should focus on recovery of lost function by directly training that function (Taylor & Broomfield, 2013). The remaining intentions suggest the use of preserved cognitive skills to teach compensatory approaches (Taylor & Broomfield, 2103). These two approaches mirror the implicit/explicit training distinction made by Klingberg (2010) that was discussed in the previous chapter.

In regards to the direct training approach, Robertson and Murre (1999) specifically addressed the issue of how such training could target the neuroplasticity changes that underpin natural recovery. While much of what they propose is based on damage to the motor and or sensory systems, they do argue that the principles should apply equally to higher level cognitive functions such as sustaining attention and executive functioning. They specify five principles of what they term “*guided recovery*” making the fundamental assumption that neural networks underpin performance and that traumatic damage will affect some parts of the neural network, but other parts will survive relatively intact. Recovery can be facilitated if activity in the functioning part of the network can be increased such that former pathways can be reactivated or new alternative pathways established.

The first principle is nonspecific stimulation. The idea here is that any activation in the network will have beneficial effects. Animal models have shown that recovery in a rich and stimulating environment has led to changes in synaptic connectivity (Will & Kelche, 1992). Thus, enriched and stimulating environments are argued to increase the likelihood of the damaged network being activated with subsequent benefits (but see Wall and Kass (1985) for evidence of maladaptive connections being formed). In terms of cognitive recovery, the implication of this principle is that having the stroke patient engage in any form of cognitive activity should have facilitative outcomes.

The second principle is bottom-up specific stimulation. As the name suggests the activation is targeted specifically at the damaged network. Constraint induced therapy (Taub et al., 1993; Taub & Wolf, 1997) described in Chapter 1 is an example of this principle because the intervention is aimed at fostering activation in the damaged motor networks. The therapeutic process in such cases involves highly repetitive movements of the damaged part of the body. In the cognitive domain the application of this principle would involve developing specific activities known to be involved in a particular damaged system, and then to institute a regime of practice using such activities. One implication of this principle is that neuropsychological tests that are used to assess cognitive impairment could themselves serve a therapeutic role if those tests were repeatedly presented. Alternatively, activities that were underpinned by the same cognitive domains could be used for training with the neuropsychological tests used solely for assessment purposes.

The third principle is top-down specific stimulation, where activity in one part of the brain, specifically the areas that govern attention, can influence activity in damaged areas outside the attentional system. As an example they cite the changes in neural plasticity that can emerge through purely mental practice of motor movements (Pascual-Leone et al., 1995). An extension of this principle to other areas besides attention would suggest that activating

any area in a cognitive network should produce enhanced activation elsewhere in the network and should thereby increase activation in the damaged portion of the network.

The fourth principle involves inhibition of competitor circuits. In Chapter 1 Trans Magnetic Stimulation was mentioned in reference to stimulating areas in the affected hemisphere and suppressing activation in the undamaged hemisphere with the intent of restoring hemispheric balance in activation. One explanation for the effectiveness of this intervention was based on the premise that suppressing activation in the undamaged hemisphere suppressed inhibitory pathways linking the undamaged and damaged areas (Meyer, Roricht, van Einsiedel, Kruggel, & Weindl, 1995; Seyal, Ro, & Rafal, 1995). Robertson and Murre (1999) suggest that while suppression of inhibition is readily apparent in focal infarcts, it is harder to detect when damage results in higher order cognitive deficits. Consequently, it is not clear how behavioural interventions could take advantage of this principle in facilitating recovery.

The final principle is arousal. The basis of this principle is the fact that higher levels of the neurotransmitters, such as noradrenaline and norepinephrine, were associated with changes in neural plasticity, perhaps through enhanced resolution of diaschisis (Feeney, 1997). Robertson and Murre (1999) argue that arousal and attentional systems are represented in different neural networks, and that the effects of arousal can have different effects to those of changes in attention. The implication of this principle is that any intervention is most likely to have maximum effect when the patient is highly alert.

When applying these principles to clinical rehabilitation Robertson and Murre (1999) make a number of additional points. Firstly, they argue that the effectiveness of such rehabilitation principles is likely to be moderated by the size of the lesion. With small lesions the natural recovery process by itself will be sufficient to produce the neural changes. With large lesions these plasticity-enhancing processes may not be effective. In the case where

there is intermediate damage they argue that the availability of adequate levels of arousal, general stimulation, suppression of inhibition and targeted bottom-up and top-down stimulation of undamaged parts of the network can facilitate recovery.

The second point is that the ability to detect the degree of residual function post-stroke is not possible using current methods of standardised neuropsychological assessment. Robertson and Murre (1999) argue that such assessment needs to be complemented by combining assessment practices with experimental methods and theories of cognitive psychology. These experimental methods can then serve as a way of providing additional insight into both the cognitive systems that have been damaged or spared and provide insight into how rehabilitation should be targeted. The Vallat et al., (2005) study is probably a good example of these points being put into practice because using Baddeley's model of working memory, which had been developed using experimental methodologies, the authors were able to localise the deficit to a particular component of the working memory systems. They were then able to tailor their intervention to facilitate recovery in this component of the working memory system.

Robertson and Murre (1999) also address the issue of when intervention should be initiated. From their perspective, the data on plasticity changes and self-repair of neural networks indicate that rehabilitation needs to be initiated as soon as medically possible after injury.

The traditional cognitive rehabilitation and plasticity rehabilitation approaches have their obvious differences, but they also share a number of common features. Both highlight the need for adequate assessments, both stress the need for substantial practice as a key component of the rehabilitation process, and both stress the need to consult the cognitive psychology literature with regards to informing the intervention process.

### **3.3.1 Associative Networks**

The rehabilitation principles described by Robertson and Murre (1999) are all predicated on partial damage to established neural networks with the five principles aimed at reviving activity in the damaged networks. In a recent review of the current understanding of the recovery process following stroke, Murphy and Corbett (2009) make a similar argument concerning the importance of networks by stating:

It has been shown that the neurons that contribute to complex functions, such as a memory trace or engram, are not localized in a single brain region but are distributed throughout the cortex. Therefore, despite its defined circuit structure, the brain functions as a spatially distributed computational machine that routes signals along multiple pathways, each capable of adapting to changes in transmission fidelity. This diffuse connectivity, together with redundancy in neuronal processing, might facilitate recovery from stroke damage. (p. 862.)

The link between biology and cognition that encapsulates activity in neural networks is most readily apparent in connectionist (parallel distributed processing) models of cognition. The basic architecture of these models is that knowledge is represented in terms of distributed networks (often with items being represented in terms of distributed sets of features) where layers of nodes in the network are linked, with either permanent or temporary associations, to nodes in other layers of the network. Learning in these models involves the simultaneous activation in sets of nodes in different layers, with the subsequent establishment of connecting pathways between the layers. Following learning, cognition is accomplished via activation of a set of nodes in one layer, which spreads along learned pathways to activate representations in other connected nodes in other layers. While initial connectionist models were loosely based on the concept of connectivity among neurons, it is only lately that connectionist models are incorporating known biological neural processes into their models (Lerner, Bentin, & Shirki, 2012).

While current connectionist models bridge the gap between biology and cognition, in the mainstream cognitive literature the idea of activation spreading along pathways to facilitate subsequent processing has a long and well established pedigree. When trying to solve the problem of how knowledge could be represented in a computer simulation, one of the earliest attempts to describe the organisation of our acquired knowledge (semantic memory) was in terms of semantic networks. Collins and Quillian (1969) structured knowledge in terms of a hierarchical network of categories, where nodes representing the most inclusive categories were at the apex of the network (e.g., animals). All the characteristics that defined such a category (such as eats, breaths, can move) were also stored at this level, with associative links from the category node to the property nodes. The next layer down in the hierarchy contained nodes representing sub-ordinate categories (e.g., birds, fish, mammals) with their attendant unique properties (such as has wings, has feathers), and the bottom layer consisted of instances of the subordinate categories (such as canary, emu) and their unique properties. All nodes in the network were linked by pathways, such that when answering a question like “*Is an emu a bird?*”, the node representing emu would be activated, activation would then spread throughout the network. The node for bird would also have been activated by the question and activation would spread throughout the network. The intersection of activation from both nodes thus became the basis for responding that in fact an emu is a bird. The other assumptions underpinning this model were that activation took time to travel between nodes and decreased in strength with increased distance between nodes.

Challenges to the Collins and Quillian (1969) model soon falsified the assumption of hierarchical organisation. In a revision, Collins and Loftus (1975) abandoned the notion of hierarchical linkages in favour of organisation based on associative strength. In this revision, nodes were still linked by pathways. The strength of association between two concepts, however, was represented by the length of the pathway connecting concepts where the higher

the associative strength between concepts the shorter the linking pathways. This adaptation of the original model was widely adopted, and part of its appeal was the extent to which spreading activation effects could be detected across different cognitive domains.

One of the most compelling sets of findings that supported spreading activation involved associative priming in lexical decision and naming tasks. In a lexical decision task, participants are presented with a letter string that is either an English word or a non-word. The participant's task is to decide as quickly as possible if the letter string is an English word or not. Thus, the decision time to identify the letter string NURSE as a word can be altered if other words are presented briefly before the letter string. For example, if the word DOCTOR is presented shortly before NURSE, the decision that NURSE is a word is significantly faster than if NURSE was preceded by BUTTER. The notion here is that when DOCTOR is presented activation spreads through its networks, to which the node for NURSE is linked. As a result the node for NURSE is partially activated such that when NURSE appears, less activation is required to conclude that the letter string is a word. In short, the presence of DOCTOR primes the associated concept of NURSE.

Spreading activation can be seen in other cognitive domains as well. In the area of episodic memory, spreading activation has figured prominently in explanations of the Deese-Roediger-McDermott (DRM) false memory paradigm (Roediger & McDermott, 1995). In this task, participants are presented with lists of words for subsequent recall where all the words are associatively related to a common item. Participants study a list of words such as *thread pin eye sewing sharp point pricked thimble haystack pain hurt injection*, which are all associates of *needle*, and to an extent are associates of each other. When given recall or recognition tests, it is the case that memory for the associates is better than for items that were studied in other lists containing unrelated words. However, it is often the case that the common associate (*needle*), which is never studied, is confidently but falsely recalled or

recognised. According to the Associative-Activation Theory (Howe, 2005), the explanation for both better recall of the associatively related items and the intrusions of the non-presented lure is that activation spreads throughout the associative network, strengthening the representations of the list items as well as the non-presented common associate. Recall involves accessing the strongly activated items at the point of recall. Given increased activation among the related items, they are more likely to be recalled than non-related items. On some occasions the activation in the non-presented lure is also assumed to be above threshold level and is also recalled as one of the list items.

The same facilitative effect of associatively related items can also be found in short-term serial recall tasks (Tehan, 2010; Tse, 2009). For example, Tehan (2010) presented participants with 6-item lists of associatively related items, and lists of unrelated items for immediate serial recall. As was the case in the long-term recall tasks, serial recall was better for the associatively related lists than the unrelated lists. Also there were occasions where the non-presented common associate was falsely recalled.

Deese-Roediger-McDermott type lists of associatively related items have also been used to prime solutions in different types of problem solving tests. Howe, Garner, Dewhurst and Ball (2010) demonstrated that false memories also prime solutions in the compound remote associate task (CRAT) (Mednick, 1962). In this task participants are presented with three words that are all linked to a third word where all three form compound words. For example, *board*, *mail*, and *magic* can all be linked to the word *black* to form compound words such as *blackboard*, *blackmail*, and *blackmagic*. Prior to completing the CRAT, the participants studied DRM-type lists where the non-presented critical lure was also the solution to CRAT. They demonstrated that adult participants solved more problems, more quickly, when primed with a false memory than when no prime was presented.

Howe, Threadgold, Norbury, Garner and Ball (2013) also used a DRM procedure to prime responses in analogical problem solving. In the primary part of the experiment, participants were presented with analogical reasoning problems of the form of, *water is to boat as road is to....* . However, prior to this component of the experiment, the solution was primed through the presentation of DRM-type lists in the experimental conditions, where the solution word was actually presented as the first word in the list (*car, truck, bus, train, vehicle, drive, jeep, ford, race, keys*) or the solution word was the non-presented critical associate of the list items (*truck, bus, train, vehicle, drive, jeep, ford, race, keys*). In the control condition the solutions were not primed. The essential outcome of the study was that both children and adults solved problems primed with false memories significantly faster than either those primed with true memories or unprimed problems.

Both the stroke literature and the mainstream cognitive literature make reference to activation spreading along a network to influence processing elsewhere. As indicated above, spreading activation among associatively related items has a substantial impact on diverse cognitive tasks. Given widespread belief that associative networks are distributed across different brain areas (Patterson, Nestor, & Rogers, 2007), it is plausible that spreading activation in a semantic network could have a beneficial effect on the recovery process following stroke. This means that associative relatedness could serve as a top-down source of activation for the damaged areas (Robertson & Murre, 1999). Consequently, incorporating associative relatedness into the cognitive tasks was one factor that impacted on the choice of cognitive tasks to use in the battery.

### **3.3.2 Working Memory and Dual Task Performance**

The second aspect of cognitive psychology that informs the choice of cognitive tasks deals with working memory research, and is addressed for a number of reasons. Firstly, at the theoretical level models of working memory have served as a theoretical basis for developing

appropriate interventions in the stroke population (Nordvik et al., 2012; Stablum et al., 2000, Vallat et al., 2005; Westerberg et al., 2007). Secondly, the working memory literature provides a well-established dual task methodology that can be utilised in most interventions. Thirdly, a number of neural-plasticity inspired training programs have been devised to increase working memory capacity (e.g., Cogmed), and as such are directly relevant to the current concerns.

As mentioned in Chapter 2 there are distinct theoretical approaches to the concept of working memory. Their origins were in a series of studies conducted by Baddeley and Hitch (1974) that addressed the question of, “What is short-term memory used for?” Their working hypothesis was that capacity limitations in this system underpinned most cognitive activity, and as such they ran parallel studies where abstract reasoning, learning and comprehension were the cognitive domains of interest. The logic of their experiments was that one of the non-controversial facts about short-term memory was that it was limited in capacity. If it was the case that short-term memory was involved in higher cognition, then by using a proportion of short-term memory capacity on one task, there would be less available capacity for the higher order task. As a result performance on the higher order task should be impaired compared to a single task condition where all capacity was available to support performance on the higher order task. In line with their expectations, they found that when asked to maintain between four to six letters in memory for later serial recall, performance on the higher order tasks was impaired relative to when no short-term memory demands were made. Their explanation for the decrement in performance involved a trade-off between storage and processing demands of the two tasks. When working memory resources had to be devoted to storage, there were decrements in processing speed. They also presented evidence that the reverse was also true, when all resources were devoted to the processing task, deficits in storage were observed. They concluded that short-term memory was at the heart of higher

order cognition, and therefore it served the function of working memory where information could be temporarily held in an available state while subsequent processing of other information could be conducted. As Unsworth and Engle put it, “Working memory is the system responsible for active maintenance and online manipulation of information over short intervals.” (Unsworth & Engle, 2007a, p. 242.)

The distinction between processing and storage is a central tenet of working memory theory and several working memory tasks have been devised that incorporate both processing and storage. For example, the operations span task (Turner & Engle, 1989) involves the storage of words for later recall while processing mathematics problems. Thus, a trial in this task involves presenting a number of mathematics problem – word pairs (e.g.,  $(2 \times 8) - 2 = 15$  *CLOUD*,  $(9/3) - 2 = 1$  *GLASS*,  $(2+6) \times 3 = 23$  *FISH*). As each pair is presented on the screen, participants are asked to indicate whether or not the provided solution to the problem is correct, and to remember the words in sequence. Once all pairs have been presented, and problems solved, the participant attempts to recall the words in the order in which they were presented. Difficulty on this task is manipulated by varying the number of problem-word pairs on a trial. The task usually starts with two pairs per trial with the number of pairs systematically increasing on subsequent trials. Those who can recall longer lists are assumed to have greater working memory capacity than those who can only recall the short lists. While the operation span task has been widely used to explore working memory capacity, a number of other complex span tasks, for example reading span (Daneman & Carpenter, 1980) and counting span (Case, Kurland, & Goldberg, 1982) tasks, have also been devised. All these tasks entail storing items for later recall while doing some form of processing (such as reading, or counting).

There are other tasks that are argued to measure working memory. Backward digit span has often been seen as a measure of working memory (Lezak, et al., 2004), although

strictly speaking distinct storage and processing components are not readily apparent. The logic here is that digits have to be maintained in memory (storage) while the order of those items is manipulated (processing). Likewise the n-back task is frequently used as a measure of working memory (Jaeggi, Buschkuel, Jonides, & Shah, 2011; Jaeggi, Buschkuel, Jonides, & Perrig, 2008; Jaeggi, Buschkuel, Shah, & Jonides, 2014; Verhaeghen, Cerella, & Basak, 2004), but it too does not have distinct memory and processing components. In this task, people are presented with a continuous string of items and are asked to respond when a presented item is the same as one that appeared in a specified number (n) of items previously. For example, in a 1-back condition, people must indicate when the current item was identical to the immediately preceding item. In a 2-back condition a participant must indicate when the current item was identical to the item presented immediately before the next to last item. This qualifies as a working memory task because items in storage have to be continually updated (processing) with a fixed number of items having to be stored depending on the task requirements (1 item stored in a 1-back task, 2 items stored in a 2-back task, etc). Different task combinations have also been utilised, for example, Jaeggi et al. (2008) combined auditory and visual 2-back tasks using a dual task methodology.

The importance of the working memory literature lies in the fact that people who are good at tasks like the operation span task (or complex span tasks in general) perform well on other cognitive tasks. Those who have high working memory capacity also do well on diverse higher order cognitive tasks like reading comprehension (Daneman & Carpenter, 1980; Turner & Engle, 1989), sight-reading music (Meinz & Hambrick, 2010), multitasking (Buhner, Konig, Prick, & Krum, 2006), regulating emotion (Kleider, Parrott, & King, 2010), and reasoning (Unsworth & Engle, 2007b). More importantly, for current purposes those who have high working memory capacity also appear more able at controlling attention towards goal-relevant information and not succumb to the effects of distraction (Kane, Conway,

Hambrick, & Engle, 2007; Unsworth & Spillers, 2010), and can be more resistant to interference on tasks like the Stroop test (Kane & Engle, 2003).

Given the evidence in Chapter 2 that stroke has an effect on working memory, attention and executive functions such as goal maintenance and resistance to interference, the current cognitive tasks concentrate heavily in these areas. Working memory is implicated already in the use of backward span in the assessment battery, but will also include an n-back task, and working memory will also be explored using a dual task methodology.

### **3.4 Cognitive Tasks**

At this point three decisions have been made: to incorporate associative priming into some tasks, to use an n-back task as a measure of working memory, and to utilise a dual task methodology to differentially tax working memory capacity. In choosing the tasks in which to embed these aspects, the constraints that applied to the selection of neuropsychological tests still applied as the tasks needed to be administered within roughly a two-minute time frame. The additional requirement was that a dual task manipulation could readily be included in each task.

Two of the tasks selected were derived from the literature reviewed on associative networks, one was derived from the working memory literature, and one involved a language-based problem solving tasks. The tasks and their basic properties are presented in Table 3.2.

Table 3.2

*Cognitive Tasks, Administration Time and Underlying Cognitive Domains of Tests used in the Battery*

Task	Administration Time	Cognitive Domain
Common Associates Test	1-2 Minutes	Executive Function (speed of processing, inhibition); Language
Immediate Serial Recall	1-2 Minutes	Short-term memory
Snap	2 Minutes	Visuo-spatial processing; Working memory
Anagram Task	1 Minute	Executive Function (problem solving); Language Processing

Associative priming was operationalised in two tasks, an adaptation of the Compound Remote Associates Test (Mednick, 1962), and an adaption of the Digit Span task where words rather than digits were the to-be-remembered stimuli. The Compound Remote Associates Test (Mednick, 1962) involves the presentation of three words that are unrelated to each other but associatively related to a third word, which when paired with each of the stimulus words forms a compound word. The task in its original form is particularly difficult and violates the time constraints for the current tasks. Consequently, the Common Associates Test (CAT), was developed along the lines of the CRAT in that three associates of a common word were presented, and the participant was to identify the common associate. The response did not need to form a compound word, and the three stimuli were all strongly associated with the target response and could be associatively related to each other. Thus, one of the stimuli was the triple *steak, beef, raw*, which are all strong associates of *meat*.

Tehan (2010) and Tse (2009) both explored associative relatedness effects in immediate serial recall. In their studies short lists containing associatively related items, or unrelated items were presented for immediate ordered recall. Both experiments established

that the associatively related words were better recalled than the unrelated words. The current task uses both related and unrelated lists, but the lists are presented in a staircase method (increasing list lengths) as is the case in digit span.

The use of a dual task methodology in all cognitive tasks allowed for the manipulation of working memory load (task difficulty). Backward digit span also provided one measure of verbal working memory. To complement these tasks a visuo-spatial n-back task was developed. The visuo-spatial stimuli were playing cards, and the distance that had to be continually updated was 1-back. Thus, the 1-back task with playing cards as the stimuli is identical to the card game *Snap*.

The final cognitive task was a speeded Anagram task. This task was developed to provide a measure of language-based problem solving. The background to forming this task is derived from a series of studies by Novick and Sherman (2003, 2008). They explored the solution of five-letter anagrams in groups of good anagram solvers and poor anagram solvers. They argued that there were two means of solution. The first involved pop-out responses where the response simply came to mind. These responses were characterised by rapid production, usually within two seconds, and unawareness on the part of the participant as to what cognitive processes had gone into generating the word. The second means of solution was invoked when pop-out did not occur. The adopted strategy was a deliberate, serial and slower rearrangement of letters to find a solution. For this method, participants could describe the process. The factors that influenced pop-out responses were the bi-gram frequencies of the letters, spelling patterns, and having a consonant as the first letter, with uncommon letters appearing in certain positions and next to other particular letters. They argued that these constraints were processed in a parallel fashion, such that when all constraints were met, the solution was rapidly produced (Novick & Sherman, 2008).

From a stroke patient's point of view, damage to the language system may not be severe enough to cause aphasia, but it could well be sufficient enough to disrupt some of the language processes involved in anagram pop-out solutions. Thus, the distribution of pop-out responses, strategic solutions, and failures might well differ between intact and impaired language processing.

The task was also added because it complemented other language tests in the battery. Thus, the reading test of the Stroop task can be considered a test of speed of access to the lexicon where there is no degradation of the input stimulus. The phonemic fluency test can be considered as a test of access to the lexicon where items have to be generated on the basis of a single letter. The Anagram task can be considered an intermediate step in that the input is a degraded version of a real word, one that can be matched to words in the lexicon to generate a correct response.

Having decided on the cognitive tasks, the final decision involved the nature of the secondary task in the dual task situations. Tables 3.1 and 3.2 outlines the primary tasks that can all be made more difficult by incorporating in a secondary task. In this instance the secondary task was a monitoring task where, at the same time as doing the primary task, they had to monitor the environment for an event that was a cue to conduct a second response. For example, in the Anagram task participants had to solve the anagram, but also indicate whether or not the anagram was a member of a particular category. The requirement to keep goals and instructions in mind while performing cognitive tasks is seen as an index of executive functioning. In some literature, however, this type of activity is seen as test of sustained attention, and in other literature this activity is seen as a test of prospective memory, which is the ability to remember to do something in the future. Using a secondary activity therefore has implications for a range of cognitive domains.

### 3.5 Task Battery

As indicated previously, the battery consisted of four neuropsychological tests and four dual task cognitive tests. The four neuropsychological tests (Stroop Colour Word Task, Digit Span, Rey Tangled Lines Test, verbal fluency) are described first and descriptions of the four cognitive tasks (Common Associates Test, Immediate Serial Recall, Snap, and Anagram Task) follow.

#### 3.5.1 Stroop Colour Word Task

The Stroop Colour Word task (Golden, 1978) is a standardised neuropsychological test that is widely used in examining cognitive impairment. It has multiple components that measure different constructs such as speed of processing, language abilities by way of reading, and resistance to inhibition.

The Stroop task is a three component task that measures speed of reading, speed of colour naming and interference between the two tasks. The Stroop task consists of three subtasks, namely Word Reading, Colour Naming, and Colour-Word naming. An interference measure can also be derived from performance on the three components.

The current versions of all three subtasks involve presenting participants with a 100-item array where items are organised in a 5 column by 20 row format on a single A4 page. In the word reading subtest the items in the array consist of the words *red*, *green* and *blue* randomly repeated throughout the list. These items are printed in black ink. The participant's task is to simply read as many words as possible in 45 seconds.

In the colour naming task the array consists of the stimulus XXXXX in all cells, but the stimulus is presented in one of three ink colours: red, green or blue. Again the different coloured XXXXXs are randomly distributed throughout the array. The participant's task is to name the colour of the ink as quickly as possible for 45 seconds.

In the colour-word naming task the stimuli involve the three words *red*, *green* and *blue* being randomly presented throughout the array, but now the words are presented in coloured ink, not in black ink. The three ink colours are red, green and blue, but the words are never presented in the same coloured ink as the name (RED is printed in green ink). The participant's task is to name the colour of the ink in each stimulus and to do so as quickly as possible for 45 seconds. Typically, participants are much slower on this subtest than either of the first two.

In all subtests, the number of words produced in 45 seconds was the primary measure. The number of errors made was also measured, but in most instances errors were detected by the participant and were corrected as per test instructions. The test manual provides test-retest reliabilities for the word reading, colour naming and colour-word naming that range between .70 and .88.

### **3.5.2 Digit Span**

Two components of the WAIS-IV digit span sub-tests (Wechsler, 2008) were adopted, namely forward span and backward span. As mentioned earlier, it is generally agreed that forward span measures passive short-term storage, but there is debate concerning backward span. It is often seen as another measure of short-term storage, but it is also considered to be a measure of working memory capacity.

In this task, the experimenter presents digit strings aurally at a rate of one digit per second. The participant's task is to repeat the digits in the same order in the forward span version of the task, or repeat them in reverse order in the backward span task. The number of digits in each trial increases in length varying from 2 digits per list to 9 digits per list in forward span, and 2 digits to 8 digits per list in the backward span task. There are typically two trials at each list length but with backward span there are four trials at list length 2. There are 16 trials in total for both backward and forward span.

Two scores can be derived from each test. The first is the total number of sequences (out of 16) that were correctly recalled (all items recalled and all in their correct position). For this measure the scores can range from 0 to 16. The other measure is the longest list length where recall is error free. This can range from 2 to 9 in forward span and 2 to 8 in backward span. The first of these scores is reported as the chosen measure of performance on this task. According to the WAIS-IV manual, the reliability for forward digit span is .82 and for backward digit span the value is .81.

### 3.5.3 Rey Tangled Lines Test.

The Rey Tangled Lines test is a little used test of visuo-spatial processing. It is a guided visual search task that requires the deployment and regulation of attention for periods of around 10 to 15 seconds (Schneider & Asarnow, 1987), and can be distinguished from cancellation, visual organisation and speed of processing tasks (Senior, Kelly, & Salzman, 1999). The stimulus page for this test is presented in Figure 3.1.

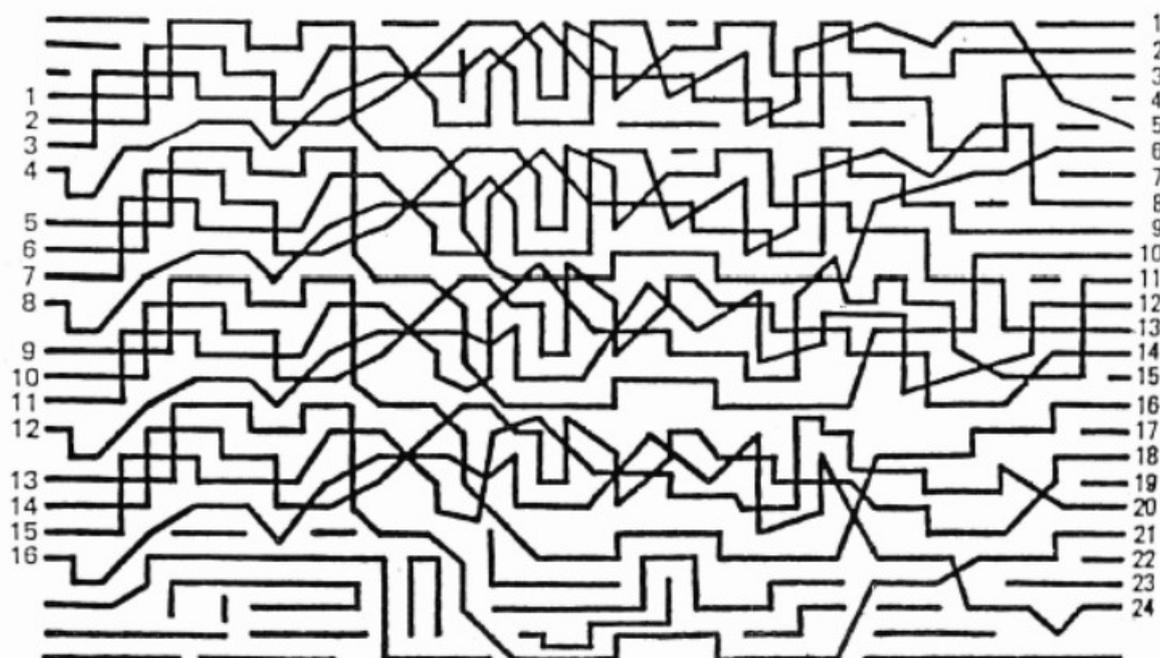


Figure 3.1. Stimulus for Rey Tangled Lines test.

As can be seen in Figure 3.1, the stimulus consists of a number of overlapping “pathways” plus additional visual clutter. Each pathway is defined by a start point on the left hand side of the figure and an endpoint on the right hand side of the figure. The starting and finishing numbers are not the same. Each trial in the task starts with the experimenter providing a start number. The participant must follow the trail from the start point to the endpoint using only their eyes and do the task as quickly as possible. The participant responds by providing the number at the end of the trail. For each trial there are two derived scores, whether or not the nominated endpoint is correct and the time taken to complete the trial.

There are two forms of the test, each consisting of six trials with different start numbers. The scores are the total number of correct solutions (range from 0 to 6) and the average completion time of the six trials. Both forms of the test show good internal consistency (alpha's of .86 and .89) and the correlation between performance on Form A and Form B is .86 (Senior et al., 1999). Form A and Form B alternated across the six sessions.

### **3.5.4 Phonemic and Semantic Verbal Fluency Tests**

There are two widely used measures of verbal fluency, the Controlled Oral Word Association Test (COWAT) (Bechtoldt, Benton, & Fogel, 1962) and the Animal Naming Test (Goodglass & Kaplan, 1972). While there are two forms of the COWAT, the FAS and the CFL, the former is the most frequently used form. The FAS version of the COWAT involves three trials. In the first trial participants are asked to say as many words as they can that start with the letter F within a 60-second time frame. The second and third trials require production of words starting with A and S, respectively. Although there are differences in the number of words produced on each trial (fewer produced with the letter A than F or S), a fluency score is created by summing the number of words produced for the three letters within the allotted timeframe. In performing this task, it is assumed that participants search

through their permanent knowledge base to access instances of each letter category, and that individuals are differentially fluent at accessing their knowledge base. FAS is assumed to reflect phonological verbal fluency by accessing their internal lexicon; that is, accessing their knowledge of words and language (Raboutet et al., 2010).

On the Animal Naming test participants recall as many animals as they can within a 60-second time frame. Animal Naming is assumed to reflect the fluency with which participants can access their meaning-based knowledge structures, often referred to as semantic memory (Raboutet et al., 2010). Semantic fluency tests have been used in a number of test batteries. Animal Naming is a component of the Western Aphasia Battery (Kertesz, 1982) and the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). Multiple categories, including animals, are used in the Set Test (Issacs & Kennie, 1973).

This description of the COWAT and Animal Naming would suggest that verbal fluency measures are non-problematic and that it is clearly understood what the verbal fluency tasks are measuring. This is not necessarily the case. There are some behavioural and psychometric indicators that suggest the two fluency tasks are measuring the same construct, yet there are other indicators that suggest different constructs are being measured (Raboutet et al., 2010). These difficulties are explored in some detail as they impact on the pattern of results that could be expected in the stroke study and they suggest additional scores to a total score.

**3.5.4.1 Similarities.** Both the phonological and semantic forms of the test have aspects in common. Behaviourally, both tasks show a rapid decline in the production rate of new items over the 60-second duration of the task. Most items are produced in the first 10 to 15 seconds of the task, with fewer items being produced in subsequent periods. Performance can be characterised as a fast-start/slow-finish across the timed interval in both instances,

even though more items are produced overall with Animal Naming than with FAS (Fernaes & Almkvist, 1998).

In both tasks, items appear to be produced in clusters (semantic clusters in Animal Naming and phonological clusters in FAS). Participants appear to generate words within a subcategory, and when items from that subcategory are exhausted or retrieval becomes difficult, the participant switches subcategories and starts again (Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998; Raboutet et al., 2010).

At the psychometric level, FAS and Animal Naming are correlated, with correlation coefficients varying from .3 to .6. In addition, studies involving factor analysis of multiple cognitive abilities tests show that both FAS and Animal Naming typically load on the same factor. Both tasks have a substantial amount of common variance suggesting that at least some of the processes involved in the tasks are identical or very similar to one another.

The similarity between tasks has led to assertions that FAS and Animal Naming are simply hard and easy forms of verbal fluency. However, an appeal to differential difficulty does not provide an explanation as to what it is about FAS that makes this task harder than Animal Naming.

**3.5.4.2 Differences.** While the FAS and Animal Naming have some similarities, there are many differences between the tasks that tend to suggest that the tasks are measuring different constructs. The mere fact that each task measures different components of one's long-term knowledge base (lexicon and semantic memory) implies that these two domains may not be identical, and that the processes that apply to one may not apply to the other.

There are at least three broad areas, where FAS and Animal Naming appear to diverge. Firstly, the two tasks are differentially sensitive to normative variables. While gender effects are generally absent in both tasks, it is clear that in non-clinical populations both tests are affected by age and education (Ostrosky-Solis, Ardila, & Rosselli, 1999;

Ostrosky-Solis, Ramirez, Lozano, Picasso, & Velez. 2004). Education has a more substantial impact on phonemic fluency than semantic fluency (Ardila, Ostrosky-Solis, Rosselli, & Gomez, 2000), and age accounts for more variance in semantic fluency than education (Tombaugh, Kozak, & Rees, 1999). The dissociative effects of age and education suggest that the two tests are not equivalent measures of the one construct.

Secondly, phonemic fluency deficits seem to be more prevalent in cases where there is damage to the frontal lobes, and semantic deficits more prevalent when damage is in the temporal lobes (Ardila et al., 2006; Jones, Laukka, & Backman, 2006). Again, the two tests appear to be measuring different aspects of verbal processing.

Lastly, verbal fluency tests are widely used in clinical situations, as both measures are sensitive to a range of neuropsychological conditions. Fluency deficits are found in frontal (Herrmann, Ehlis, & Fallgatter, 2003) and temporal lobe pathology, Parkinson's Disease (Donovan, Siegert, & McDowall, 1999), schizophrenia (Chen, Chen, Chan, Lam, & Lie-Mak, 2000), subcortical dementia (Testa et al., 1998), Multiple Sclerosis, Huntington's disease (Ho et al., 2002), depression (Crowe, 1992), vascular and degenerative dementias (Cooper et al., 2001), and traumatic brain injury (Axelrod, Tomer, Fisher, & Aharon-Peretz, 2001).

Although sensitive to these conditions, it is not uncommon for clinical cases that show a large deficit on COWAT scores to have unimpaired performance on Animal Naming or vice versa. For example, those with Huntington's disease show a deficit in FAS scores but not with Animal Naming compared to a control group (Larsson, Almkvist, Luszcz, & Wahlin, 2008). The reverse was true with those suffering from Alzheimer's disease with a deficit in category fluency but not letter fluency (Fernaesus, Ostberg, Hellstrom, & Whalund, 2008). Moreover, it is possible to find instances where opposite dissociations are found within the same population. Thus, Zec et al. (1999) found an animal naming deficit with Parkinson's patients but no FAS deficit. In contrast, Azuma, Cruz, Bayles, Tomoeda and

Montgomery (2003) found deficits with both semantic and phonemic fluency in a Parkinson's group, but the phonological performance deteriorated more rapidly across a 2-year period.

These findings suggest that there is not a simple explanation for the similarities and differences between the two forms of verbal fluency.

When trying to resolve the issue of similarities and differences between the two tasks, most researchers argue that different internal knowledge bases are accessed during FAS and Animal Naming tasks. However, it is quite possible that the access or retrieval processes involved in both tasks might be identical in each case, which could account for some of the common variance that underpins the correlations between the two tasks. There has been some research exploring the time course of responses during the production period and two distinct processes have been identified that appear to be used at different stages in the response period.

**3.5.4.3 The two processes approach.** Fernaeus and Almkvist (1998) report a study in which 126 patients who were referred to a memory clinic for suspected cognitive impairment underwent a neuropsychological examination including the COWAT test. In examining performance, the 60-second response period was divided into six 10-second sections. The scores for each period were entered into an exploratory factor analysis resulting in two factors. The scores in the initial phases (10-sec, 20-sec and 30-sec) of each trial loaded onto the first factor. The second factor was characterised by high loadings from the latter phases of the response period. The fact that two factors emerged was taken as evidence that two distinct processes were operating across the response period, one impacting early in response generation and the other having more of an impact later in the response period. In order to establish what these processes were, they conducted a second factor analysis that included factor scores for the early and late periods, scores from the Information, Similarities and Digit Span subtests of the WAIS-R, a vocabulary test, and a free recall test. Again, two factors

emerged. The FAS early phase variable loaded on the same factor as digit forward, digit backward, and free recall. The authors argued that this factor represented semi-automatic retrieval, and further argued that this represented rapid and semiautomatic retrieval from the lexicon. The second factor was characterised by loading of the FAS late phase, Information, Similarities and Vocabulary variables. This factor was interpreted as an effortful retrieval factor that was characterised by slow and effortful retrieval from the lexicon. Results of this experiment indicated that through the 60-second response period there were two distinct retrieval modes that contributed differently to the initial and later response periods of each trial of the FAS test.

The clinical aspects of production differences across the various phases of the response period have been noted recently. Bittner and Crowe (2006) examined the effects of Traumatic Brain Injury (TBI) on the FAS. In their experiment, the TBI group comprised a portion of participants who had a naming difficulty and a group who had no such difficulty. These participants' performance was compared to that of a non-impaired control group. They examined the 60-second response period in four 15-second epochs and found that difference between TBI and control groups could be localised to the first 15-second time period. Moreover, there were no differences between the two clinical subgroups.

In a similar study, Larsson, Almkvist, Luszcz and Wahlin (2008) explored verbal fluency in a group who carried the gene for Huntington's disease, but were not showing any symptoms. Their performance was compared to a group of participants who did not carry the gene. Both phonological and semantic fluency were tested and again performance was examined across the response period, in this instance across six 10-second time slots. The results indicated that there was no difference between the groups on semantic fluency measures, but there was a deficit on phonemic fluency. Again, this deficit was most pronounced in the initial phases of production. In addition, the different fluency measure

correlated with other tests; semantic fluency correlating with crystallised measures of cognitive ability and phonemic fluency correlating with speed of processing and working memory.

Ferneaus et al. (2008) compared production of participants with Alzheimer's disease to those reporting Mild Cognitive Impairment over 10-second intervals using the standard 60-second period. They found that the two groups could be discriminated on the basis of performance on the first 30 seconds of the trial. Raboutet et al. (2010) explored clustering and switching effects over the time course of the response period. They too found evidence for rapid and automatic access of semantic memory during the initial portion of the response period, and noted that the degree of switching, and to a lesser extent clustering, differed across the different phases of the response period.

Consequently, the literature is converging on the notion that there are at least two distinct processes going on during response period. One process is relatively fast and automatic, and has its influence in the early stages of production. The second is slower and more strategic in nature, and has its impact later in the response period. The clinical data tend to show that the first phase of the response period is the one that is most sensitive to cognitive impairments.

One further problem relates to psychometric issues. The reliabilities are generally much higher in the phonemic fluency task than the semantic fluency task. This is due to the fact that the phonemic fluency task has three trials (F, A and S) whereas the semantic fluency task only has one trial (Animal Naming). Because there is only one trial in the semantic condition, two further categories have been added to the semantic fluency task used in the battery to improve the test's reliability, and at least to make it formally equivalent to FAS. Countries, and Fruit and Vegetables were selected for the additional trials as they have been used in other studies involving verbal fluency (Ostrosky-Solis et al., 2004).

**3.5.4.4. Phonemic fluency.** The three trials (FAS) of the COWAT served as the phonemic fluency task. On each trial a letter was presented to the participant and they had 60 seconds in which to produce as many words beginning with that letter as possible. The experimenter recorded the number of words produced in each of the four fifteen-second intervals. The number of words produced in the first 15-second segment served as the score for 1<sup>st</sup> responses. The number of words produced in the 45 seconds that constituted the second, third and fourth 15-second segments served as the score for later responses. The total score, which is the standard measure for the task, was the sum of the 1<sup>st</sup> and later response scores.

**3.5.4.5 Semantic fluency.** Three trials using Animal Names, Countries, and Fruit and Vegetables as the stimuli constituted the semantic fluency task. Participants were given the category label and they had 60 seconds in which to produce as many members of the category as possible. Scores were generated for 1<sup>st</sup> responses, later responses and total scores.

### **3.5.5 Common Associates Test**

In searching for cognitive tests that involve the presentation and subsequent interplay of associative related items the one test that did emerge was the Compound Remote Associates Test (Mednick, 1962). On each trial in the test three items are presented that have no apparent relationship to each other but are all associatively related to a fourth word in that when combined to that fourth word they form compound words. For example, *night*, *stop*, and *wrist* are part of the compound words *night watch*, *stop watch* and *wrist watch*. This test was originally designed to measure creativity (Mednick, 1962), but it has also been used as a measure of problem solving ability (Howe, Garner, & Patel, 2013). Bowden and Jung-Beeman (2003) explored solution time to these problems based on time limits of 2 seconds, 7 seconds, 15 seconds and 30 seconds per problem. More items were solved given longer study periods, but even with 30 seconds per triple only half the problems were solved. In short, the

Compound Remote Associates Test is difficult for young healthy participants. However, while the task meets the needs of exploring associative relationships, it is likely that the task itself is too difficult for a stroke population in that performance might well be on floor.

The basic ideas behind the Compound Remote Associates Test spurred the development of a new, easier task called the Common Associates Test. As in the Remote Associates Test, each test stimulus consists of three words that are unrelated to each other but are all related to a fourth word. In this case the associative relationship is much stronger than the Compound Remote Associates Test and the solutions do not have to be compound words.

**3.5.5.1 Primary task.** The Nelson, McEvoy and Schrieber (1998) free association norms served as the basis for the development of the Common Associates Test (Common Associates Test). The Nelson et al. norms were created by presenting a large group of participants with a cue word and asking the participants to write down the first word that came to mind in response to the cue. In this procedure a number of different items are produced, but many participants will give the same response as other participants. Associative strength is represented by the percentage of participants who produced the same response. For example, in the case where *hot* was the cue word, 68% of participants responded with *cold*, 9% said *tub*, 3% said *warm*, and so on.

To create the Common Associates Test, 150 four-word sequences were initially selected from the norms. One of the words was the target word and the three other cue words were associatively related to the target word. For example, if the target word was *horse*, three associative cues were *pony* (*pony-horse* associative strength = .75), *saddle* (*saddle-horse* strength = .88), and *gallop* (*gallop-horse* strength = .66). In choosing the three cue words an attempt was made to select strongly related cues with the result that the average cue-target strength (technically backward associative strength) for all cue-target pairs was .44 (SD = .12).

Selecting 150 such four word sequences with strong associates was quite difficult and in order to incorporate dual task components and have sufficient trials in each test, only three different versions of the test were created. To create the three tests, the four-word sequences were rank ordered in terms of the average associative strength. The items were then sequentially allocated to the three tests such that the average associative strength (and standard deviation) of the 50 sequences making up Forms A, B and C of the test were the same as the parent population. All three forms of the test were thus matched for pre-existing associative strength.

In order to examine dual task effects, each of the three forms was then divided into two parts (A for event-based and B for time-based), with the first consisting of 20 trials and the second consisting of 30 items. In each form the words were presented in either lower case, uppercase, underlined or in italics. While the majority of the words were in lower case, on any one trial the three words could involve any combination of the four formats. The use of different formats formed the basis of monitoring the secondary task in that they could monitor for underlined words, words in italics, or words in upper case letters.

*Instructions and presentation.* The instructions explained the nature of the free association procedure used to generate the stimuli, and described the relationship between the three items stressing that all three items were strongly associated with the one word. They were instructed to try and identify the common word. They were told that there was no single correct answer and that they were to output the first word that came to mind.

The test was designed to be presented on either a computer using Microsoft Powerpoint software. On each trial the three cue words were presented in the centre of the screen in Calibri 48 font for 6 seconds. Participants were instructed to produce the first word that came to mind or to say pass and go on to the next triple. The number of targets produced was the primary measure on this task.

**3.5.5.2 Dual task component.** The variation in presentation format served as the cue for both event-based and time-based cues. Event-based responses were sought in Part A of the test and time-based responses were sought in Part B.

*Event:* Part A of each form consisted of 20 three-word trials, and on eight of these trials one of the words on the list was presented in upper case. The eight trials were randomly ordered with the other remaining 12 trials. Participants were told that any time one of the list items was presented in capital letters they were to say “Capitals” after they had provided the common associate.

*Time:* Part B of each form consisted of 30 three-word trials. On nine of the trials one of the words on the list was presented in upper case, and on nine of the trials one of the words was underlined. The instructions indicated that they now had to monitor the input for two types of words, those in capital letters and those that were underlined. They were also asked to monitor a digital clock and that they had to switch from monitoring one type of word to the other every 30 seconds. They were told that they should start with monitoring words in capital letters, then after 30 seconds they were to switch to monitoring for underlined words, and after the next 30 seconds switch back to monitoring for words in capital letters, and so on. Given the presentation timing characteristics of 6 seconds per triple, the 30-second time period resulted in a switch after every five trials. These timing parameters produced five opportunities for the participant to correctly switch from one task to the other. In each group of five trials, two of the trials consisted of target items and one trial contained a distractor, such that if the switches were done correctly there were 12 opportunities for the participant to make a correct response to the targets.

**3.5.5.3 Scoring.** The score on the primary task was simply the number of common associates produced. In Part A of each test the maximum number was 20, and in Part B the maximum number was 30. In the event-based component there were eight opportunities to

correctly identify the word in capital letters for Part A. Therefore the maximum event score was eight. In the time-based component, there were nine opportunities to identify a word in capital letters and nine opportunities to identify an underlined word. However, given the structure of Part B of each test, if participants switched at the right time on all occasions, then there were 12 opportunities to correctly identify the word. Therefore the maximum score for the events component was 12. There were five opportunities to make a correct switch from one task to the other, so consequently there was a maximum score of five for the time-based component of the task.

### **3.5.6 Immediate Serial Recall**

This task is based on the forward Digit Span task that is used in most standardised tests of cognitive abilities. Digit span in the neuropsychological testing domain is seen as a measure of attention and concentration or is incorporated into measures of working memory. In the cognition literature digit span is seen as a measure of short-term memory rather than working memory because digit span only requires the passive storage of information, not storage and additional processing that is characteristic of working memory tasks. Thus the task was included as an index of short-term memory capacity.

Rather than use digit span per se, words were used as the memory material in the current test. The choice of words was based on the fact that associative relatedness could be incorporated into the task (Tehan, 2010; Tse, 2009).

**3.5.6.1 Primary task.** In this task the experimenter reads aloud short lists of words, and the participant attempts to verbally recall the items in the order in which they had been presented. Typically, the number of items in each list is low to begin with and gradually increases to make the task increasingly difficult. On each version of the test the starting list length was two words, and memory was tested at this length for four trials. On the fifth trial the list length increased to three items, and this length was maintained for another four trials.

Subsequently, trials of four, five and six items were tested with four trials at each list length. The combination of five list lengths and four trials at each list length meant that participants were exposed to 20 trials in total.

There were two types of trials at each list length. In one type of trial all the items in the list were common associates of a word (that was never presented) and in many instances these words were related to each other. For example, *bed rest awake tired dream wake* are all associates of *sleep*. In the other type of trial the words were not obviously related to each other (*pistol hate yard dam peach grain*). The associatively related items were derived from the Tehan (2010) study which in turn was taken from norms used to study false memory effects in long-term memory (Stadler, Roediger, & McDermott, 1999). In these norms, the associates of a lure are provided and the 10 sets that produced the greatest number of false memories were selected. The six items that were most strongly related to the lure were chosen as the stimuli for the experiment. In the associates list, the items were always presented in order of strength with the strongest being presented first. The unrelated items were taken from a study by Tehan (2010), which showed that immediate serial recall was better for the associatively related lists than the unrelated lists. In the unrelated lists the items were always presented in the same order, as was the case with the associatively related lists.

Initially, six different versions of the ISR task were generated. Within each block of four trials, the order of the related and unrelated lists was randomised, but the same order was used across the six different versions of the task. What varied across the different versions was the assignment of list to list length. The sleep related words, therefore, might have been assigned to a five-item list in one test, but assigned to a three-item list on another test.

*Instructions and presentation.* The instructions were those typically used in digit span. Participants were told that the experimenter would read out a short list of words and that their

task was to repeat back the words in the order that the experimenter had spoken them. The words in each list were read in a monotone voice at a rate of about one word per second.

**3.5.6.2 Dual task component.** The cue for both single-event and two-event monitoring tasks was taxonomic category membership.

*Single Event.* Two of the tests were designed to be used in a single task configuration. The remaining four versions were altered to incorporate a secondary task. At each list length one of the items on each dissimilar list was replaced by a member of a taxonomic category that was unrelated to any of the other items in that list. For example, one of the trials was *nail black smell tennis jazz*, with *black* being an instance of the category *colours*. Each of the unrelated trials contained an instance of the category, and the target word was always presented among the first three items in the list irrespective of the list length that was being tested. The two categories were *an item of clothing*, and *a colour*.

The instructions on these versions of the test explained the dual task nature of the current version of the immediate serial recall test. Participants were told that some trials would contain members of a particular category, and that their task was to first recall the items on the list in order and then if one of the words was from the nominated category, they were to mention what the member of the category was. In the example above, participants would say “black” after they had attempted to recall the five words in order.

*Two-events.* The two two-event versions of the test were implemented by including items from two different categories in each of the unrelated lists as per the single-event tests. Again the target items were both presented among the first three items in the list at all list lengths. The categories were *types of metal* and *wild animals*, and *liquids* and *types of fish*. Participants were again told that they would have to monitor the trials for instances of these categories and to report them after they had recalled the list items in order.

**3.5.6.3 Scoring.** Several scores can be derived from the Immediate Serial Recall task. Given that the task is based on the Digit Span task, one could generate the same score by totalling the number of sequences recalled correctly. Alternatively the number of items correctly recalled in position can be summed across the different list lengths. This latter measure was adopted as it was more sensitive, giving partial credit on lists where some but not all items were recalled correctly. Consequently, the possible total score was 40 both the associatively related lists and the unrelated lists. In the event-based tasks the maximum score was 10 items in the single-event and 20 in the two-event condition.

### **3.5.7 The Snap Task**

While snap is a card game that is commonly played by children, computerised versions of the task are found in commercial products such as CogState ([cogstate.com](http://cogstate.com)) and CogSport ([www.axonsports.com](http://www.axonsports.com)). The card game is a specific instance of the more general n-back task that is commonly used to measure attention/working memory performance. In the n-back task people are presented with a continuous stream of events that they must continually monitor and update. The primary task is to detect item repetitions. The playing card version of the n-back task was used to provide an instance where pictorial stimuli were used rather than words or verbal stimuli, which figure heavily in a number of the other tasks in the battery.

**3.5.7.1 Primary task.** Six different versions of the snap test were created in which each test involved the presentation of 104 playing cards (two packs of cards) in which five “pairs” were embedded. Each “pair” consisted of two cards in succession sharing the same number or face. For each test, the 104 cards were first randomly ordered with any “pairs” that were generated in the randomisation process being removed by swapping nearby cards. The five pairs were then created by swapping cards such that the pairs occurred in positions 15 and 16, 26 and 27, 52 and 53, 77 and 78, and 86 and 87. The five pairs in each test consisted

of different numbers or face cards. The choice of these positions was to control attention and expectations over the early part of the test. The first two intervals, therefore, were fairly close with the third and fourth being more widely separated and the fifth going back to a short interval.

*Instructions and presentation.* The 104 cards on each test were assembled into a slideshow using the Microsoft Powerpoint program, with each card being stored on a separate slide. Participants were told that they were going to play the children's card game called *Snap*. They were told that hitting the space bar on the laptop keyboard would "flip" a card. Each flip removed the existing card from the display and presented the next slide. For each card they "flipped" they were asked to indicate by saying YES or NO as quickly as possible if the current card had the same number or royal face as the previously presented card. The requirement to make a response on each card is not typical of the snap game but was required here as a means of monitoring the participant's attention. They were asked to perform this task as quickly as possible and time to complete the task was monitored.

**3.5.7.2 Dual task component.** Single-event monitoring components were added to the task in two versions of the test, and two-event components were added in two versions. In all cases participants had to monitor and report any time a card with a specific number or face appeared, in addition to the pair matching task.

*Single-Event.* The event-based task required the participants to monitor the cards with a specific number (e.g., a 7) or a specific face (e.g., a queen). Given the constraints of a deck of playing cards, there were eight events possible in each test. The event card was never one of the "pairs". The event card always appeared in fixed positions within the test with roughly 12 cards intervening between events. These positions were 12, 23, 35, 47, 59, 72, 83 and 97.

The instructions for these tests indicated that participants were still to do the Snap task as per usual, but this time they had to complete a second task at the same time. They were

told that they had to monitor the cards for a specific number or face and that when the card appeared they were to first make the yes or no decision regarding its identity to the previous card and then state the number or picture required. If the card they were monitoring was a queen and a queen appeared on the screen, they were to say yes or no and then to say “queen”.

*Two-Event.* The two-event conditions were implemented by including a second card that also needed to be monitored in addition to the pair decisions. Again there were eight presentations of this second to-be-monitored card on each test and the card was always presented at the same position across tests. The second to-be-monitored card always appeared at positions 11, 22, 33, 46, 58, 71, 82 and 96. The instructions for these tests indicated that participants were still to do the Snap task as per usual, but this time they had to look for two different cards.

**3.5.7.4 Scoring.** The participant’s use of the space-bar means that they had control over the presentation rate of the items on the test and the instructions stressed that the task needed to be done as quickly as possible. The first measure taken was the duration of the test. Performance on the primary task was also measured by counting the number of times the yes response was given to the pairs. A total score of five was possible. For the single-event task, the score, out of a possible total of eight, was the number of times to-be-monitored card was correctly identified. The total possible score for the two-event condition was 16

### **3.5.8 The Anagram Task**

The planned battery had a number of tasks where ability to process words was examined. The easiest task involved reading words in the read component of the Stroop task. At the more difficult end, in the Phonemic Fluency task, participants must generate words on the basis of a single letter. The idea of the Anagram task was based on the need for a task that

was not quite as automatic as reading, but not quite as difficult as generating items from a single letter. Solving anagrams where a visually presented word has two letters transposed is similar to reading in that often the word looks as though a typographical error has been made, but at the same time one still has to generate an item to a letter string rather than a single word. The task is seen to be a measure of the ease with which people can utilise the lexical representations that underpin performance in reading and lexical access tasks (Fink & Weisberg, 1981; Muncer & Knight, 2011), and is therefore highly related to language abilities. The task was also devised because it was amenable to dual task manipulations.

**3.5.8.1 Primary task.** The anagram solving task involves presenting participants with an original word, where two internal letters have been transposed so that the word is no longer intact. The task requires participants to mentally un-jumble the letters and then to identify the original word. Task difficulty has been manipulated by varying the number of letters in the word. For example, in a seven-letter word, transposing two letters leaves a greater proportion of the word intact (*statoin*) than transposing two letters in a five letter word (*spaer*). This is the process that has been adopted to create the six different 80-item anagram tests.

The items in the anagram test were 398 concrete words that were selected from the MRC psycholinguistic database ([http://www.psy.uwa.edu.au/mrcdatabase/uwa\\_mrc.htm](http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm)). A minimum concreteness rating of 500 was used as the inclusion criterion. The word pool consisted of 77 seven-letter words, 144 six-letter words, and 177 five-letter words. The majority of the anagrams were created by swapping two internal letters in each word, with the preference being to swap letters towards the end of the word. In the case of a five-letter item the third and fourth letters were swapped, with six- and seven-letter words the fourth and fifth letters were swapped. This meant that the first two or three letters remained intact and the final letter was maintained in its normal position. For some words where reversal of the

above mentioned letters would result in either an identical (e.g., *troop*) or alternative (e.g., *trial*) word, other letters were reversed.

The structure of all six tests consisted of 80 anagrams that were presented in upper case in four blocks of 20 anagrams, although the block structure was transparent to the participant. The first 20 items on each test were seven-letter strings, the next 20 were six-letter strings, and the third and fourth blocks consisted of five-letter strings. The items were randomly assigned without replacement to each of the six tests, and given the number of items in each pool, three unique tests were created before items had to be repeated. This meant that on the fourth test, some of the seven-letter strings repeated from one of the earlier three tests. By test five the five-item strings started to be repeated. The 80 items on each test were prepared as Powerpoint slides with each word centred in size 48 Calibri font.

*Instructions and presentation.* The instructions explained what an anagram was. Participants were told that each time they pressed the space bar on the laptop an anagram would appear, and that they had to simply say what the real word was if they were able to solve the anagram. Instructions asked for the participants to work as quickly as possible and not spend too much time on any word they found difficult. In the case of difficult items they were to say “pass” and move on to the next word.

**3.5.8.2 Dual task component.** Single-event monitoring components were in two of the tests, and two-event components were added in two of the tests. The cue in all tasks was category membership. Ten instances with six or more letters from six different taxonomic categories were selected from the University of South Florida category norms (McEvoy & Nelson, 1982). An anagram of each instance was created using the principles described above.

*Single Event.* On each of the single-event tests the 10 target items were presented in the first three blocks. Four were presented among the seven-letter strings, three were

presented among the six-letter strings, and three were embedded among the five-letter strings. Specifically, the targets appeared at positions 3, 8, 14 and 20 in the first block, positions 4, 10, and 17 in the second block, and positions 3, 8 and 11 in the third block. The categories were *a fruit*, and *a country*.

The instructions on these versions of the test explained the dual nature of the task. Participants were told that some trials would contain members of a particular category, and that their task was to first solve each anagram. If one of the words was from the nominated category, they were to mention what the category was. Therefore, if the letter string in the fourth test was *ORAGNE*, they were to first solve the anagram by saying “orange” and then say “fruit”.

*Two-event.* Two-events were implemented by including items from two different categories among the letter strings as per the single-event tests. The item from the second category appeared immediately in the list after the first target. The categories were *animals* and *body parts*, and *vegetables* and *birds*.

**3.5.8.3 Scoring.** Two scores were taken on the primary Anagram task, the number of correct solutions and the time it took to complete the task. In the single-event condition tasks the maximum score on the secondary task was 10 items, and 20 items on the two-event tasks.

### 3.6 Testing Protocol

Having decided what tasks to use, decisions needed to be made concerning how frequently and under what conditions the testing was going to take place. Again, the cognitive training literature was considered to determine how learning, particularly on the cognitive tasks, might be maximised.

With regards to procedural aspects of how to conduct training studies, there are a number of attempts to define the characteristics of a successful intervention. Klingberg (2010) argues for three principles: training should not be focused on strategies but rather for

plasticity changes; training should be focused only on working memory tasks to enhance plasticity changes; and training schedules should be intensive with constantly adapting levels of cognitive load to ensure that the training is always challenging. Klingberg suggests that about 20 hours of training (20 sessions of 1 hour duration) is the minimum required.

In a similar vein, Morrison and Chein (2011) noted that there is a degree of commonality among the procedures used in training programs. At a global level all involved the repetition of cognitively demanding tasks targeting domain-general working memory capacity. More specifically they mentioned eight different characteristics: limiting strategy use; minimising automatic processing; employing multiple modalities; involve sequential processing and updating; maintaining information in the face of interference; enforce rapid encoding and retrieval; can be adapted to cater for individual differences in ability; and can be adapted to maintain high cognitive workloads and intensive engagement.

Morrison and Chein (2011) also noted that there are some differences among training programs. Some programs (e.g., Cogmed, COGITO) train on multiple tasks that address multiple components, increasing the likelihood that some tasks will generate training gains. However, such a multi-faceted approach means that the specific locus of any training gain is hard to determine. Other training programs utilise a single task (Verhaeghen et al., 2004) making it harder to find a training benefit, but easier to identify the cognitive mechanisms involved. Other researchers use a variety of tasks that address one component of working memory. For instance Dahlin and colleagues (Dahlin, Neeley et al., 2008; Dahlin, Nyberg et al., 2008) use a variety of tasks that involve working memory updating.

These training characteristics informed some of the decisions in constructing the testing protocol. The tasks involved both verbal and visuospatial tasks, and they required rapid encoding and retrieval. Building in two levels of dual task performance meant that levels of difficulty could be varied by going from single task to dual task with one event, and

subsequently to dual task with two events. In addition with many of the tasks being time pressured, there was inbuilt motivation to maintain active engagement with the task at all times.

The protocol did differ in three respects from the recommended pattern. Firstly, the task battery involved more than just working memory tasks. Examining diverse aspects of cognitive performance, together with the multi-faceted nature of each task, means that emergent repetition effects might be difficult to interpret in terms of which specific components of cognition had improved. Secondly, the length of each session was limited to approximately 40 minutes to control for fatigue, illness and problems in concentration (Lezak et al., 2004; Nys et al., 2005a). Thirdly, because the training was taking place in a hospital ward, there was never any realistic likelihood that 20 one-hour training sessions could be conducted. There is increasing pressure in Australian hospitals for patients to be discharged from hospital as quickly as possible. The pilot testing in the hospital setting indicated that in all likelihood six sessions would be a maximum before the patient was discharged. Thus, the decision was made to administer two sessions of single task training, two sessions of single-event dual task training and two session of two-event dual task training.

### **3.7 Chapter Summary**

The chapter started by examining the constraints that had to be addressed in selecting a test battery to examine changes in cognitive performance in the early weeks following a stroke. The development of the battery was informed by the prevalence and training studies outlined in Chapter 2, and by the cognitive rehabilitation literature (Ponsford et al., 2013; Wilson, 2009). For current purposes there were a number of factors that influenced the development of the battery.

The first was the need to assess for cognitive impairment. In response to this step specific neuropsychological tests that addressed the nominated cognitive domains were

selected such that the tests could a) determine which cognitive processes were impaired by the stroke and which remained intact, and b) be used to track improvement or lack thereof over the course of a patient's stay in hospital following the stroke.

The second step was to consult models of cognitive functioning in determining appropriate cognitive tasks. To this end, the literature exploring training for neural plasticity was first consulted. Two aspects of cognitive psychology were also deemed to be relevant, one dealing with spreading activation in established semantic networks, and the other dealing with dual task methodologies employed in the study of working memory capacity.

Detailed information about the four neuropsychological tasks and the four cognitive tasks was then provided. These tasks represent the battery that forms the core of the empirical component of the thesis.

The third factor was to detail how the program was to be delivered. Decisions in this regard were informed by recent literature that has examined working memory training effectiveness. While a number of the characteristics were incorporated into the testing protocol, pilot testing indicated that six sessions were likely to be near the maximum number of sessions that could be held between the period when the patient was deemed medically stable and when they were discharged from hospital.

In sum, the planned battery consisted of four standardised neuropsychological tests (i.e., Stroop, digit span, Rey Tangled Lines, and verbal fluency tests). These tests have been the subject of much research, and their psychometric characteristics and the effects of different demographic variables on performance have been well documented. Moreover, large-scale representative samples have been used to create normative data on which individual performance can be evaluated. However, it is the case that changes in test performance with repeated administration over a short time frame have not been established as yet.

The cognitive tests have not been standardised, although for some tests there is a reasonable literature base to support their use (i.e., n-back, immediate serial recall, and Anagram tasks). The psychometric properties of the tests have not been established, and there is no normative data available to make comparisons at the individual level. Like the standardised tests, there is also no existing data concerning changes in performance with repeated testing.

In the next two chapters the properties of the tests are explored. In Chapter 4 the cognitive tests are examined using a young sample with a view to their applicability to the stroke population. In Chapter 5 the full battery is administered to a sample of older participants that cover the age range where 85% of strokes are observed.

## **Chapter 4: Evaluation of the Cognitive Tasks**

### **4.1 Chapter Overview**

Chapter 3 outlined four cognitive tasks that were developed for a new battery. One of the developed tasks was the Common Associates Test, which is conceptually related to the Compound Remote Associates Test (Mednick, 1962). Having been developed specifically for the current research, it has not been the subject of any empirical analysis. The other tests, snap, immediate serial recall and the Anagram task, are all specific adaptations of existing tasks that have a large experimental literature from which to draw expectations. This chapter begins the process of evaluating the four cognitive tests in the battery.

The battery has to meet a number of criteria to be useful. Since the primary thrust of the research involves the changes in behaviour across testing sessions, any factor that might compromise the chance of observing change or the interpretation of such change needs to be addressed. The first issue is the reliability of the test. While there are a number of measures of reliability, alternate-forms reliability for the Common Associates Test and test-retest reliability, for the remaining three tasks, are of prime importance. Both forms of reliability provide an index of the degree to which the result on the first test occasion will match those on the second testing. Since all tests have some error associated with performance, results across two testing sessions are rarely going to be identical. However, the more reliable the test is the less variation in scores there will be with repeated testing (Strauss et al., 2006).

Test-retest reliability is important in present circumstances because it has a bearing on the interpretation of any increases in performance across testing sessions. Detecting improved performance across testing sessions is more likely to be observed when reliability is high than when reliability is low (Strauss et al., 2006). Essentially, only large improvements can be detected with tests that show lower levels of reliability. In evaluating the reliability of the

measures two sets of results are presented: The bivariate correlations between all testing sessions, and the intraclass correlations between successive testing sessions. Calculating the intraclass correlations was based on a two-way mixed consistency model. Shrout and Fleiss (1979) provide the following guidelines for assessing the obtained correlation. Intraclass correlations of 0–.20 were viewed as slight, .21–.40 as fair, .41–.60 as moderate, .61–.80 as substantial, and over .81 as excellent agreement. There has been debate concerning the precise values of what constitutes acceptable reliability (Lexell & Downham, 2005), but estimates of .70 and above are generally seen as representing good reliability (Calamia, Markon, & Tranel, 2013; Lexell & Downham, 2005; Strauss et al., 2006). Consequently, if the intraclass correlations between successive testing sessions were consistently above .70, the task was considered to have acceptable reliability. In the case where reliability was low in the early sessions but improved with repeated testing, such tasks were also considered to be acceptable. However, in cases where the tasks showed correlations of below .70 on most test sessions, those tasks were not considered to meet the reliability criterion.

Measurement error is one source of error that has an influence on the likelihood of observing systematic changes in performance across sessions. A second source of variability that could compromise interpretation of changes in performance involves the equivalence of different forms of any particular test. In a number of cognitive tests where tests need to be repeated, alternate forms are often employed to guard against specific practice effects. In the case where alternate forms are used in tracking changes in performance across training, any differences that exist between the forms can be problematic in that improved performance could be confounded with changes in the difficulty on each of the forms of the test. Such interpretive problems could be avoided if the alternate forms were all equally easy or difficult. Thus, in the instance where there are alternate forms of a test, the equivalence of those forms is a desired attribute (Rogers, 2010). In addition to performing analyses of

variance, intraclass correlations involving successive administrations of the different forms were conducted to determine alternate-form reliability estimates.

A third criterion for evaluating training effectiveness is the degree to which performance on the task improves with repetition. The training literature with cognitively intact participants shows that performance on most training tasks improves with increasing levels of practice (Morrison & Chein, 2012, Shipstead et al., 2012). In the training literature, the absence of improvement would be problematic. In the case of neuropsychological tests it has been well documented that on many tests performance increases on a second testing. In the absence of any changes in the underpinning cognitive abilities such improvements have been label as practice effects. Such practice effects have traditionally been seen to introduce unwanted noise into measurement and thus need to be controlled (Strauss et al, 2006). However, as a number of authors have recently suggested the presence, absence or differential strength of practice effects have potential to provide clinically useful information (Darby et al., 2002; Duff, Callister, Dennett, & Tometich, 2012). For example, Darby et al. (2002) administered a short computerised test battery four times in the space of 3 hours to a group of older patients who had mild cognitive impairments and to a group of healthy matched controls. The control group showed substantial improvement on the battery across the four sessions. In contrast the clinical group showed severely attenuated learning across the four sessions. The authors concluded that the size of the practice effect was a good marker of cognitive impairment.

In exploring practice effects in non-stroke populations, the strength of practice effects is negatively correlated with cognitive impairments (Calero & Navarro, 2004; Duff et al., 2011), and is positively correlated with long-term cognitive outcomes (Duff et al., 2011). It is also associated with the beneficial effects of training interventions (Calero & Navarro, 2007; Duff, Beglinger, Moser, Schultz, & Paulsen, 2010). A recent meta-analytic study of the

practice effect has shown that the strength of practice effects is not identical for each task, and are moderated by age and test-retest intervals in a negatively correlated fashion (Calamia, Markon, & Tranel, 2012).

In light of the research exploring practice effects, the use of repeated testing for exploring behaviour changes in patients with stroke needs to consider the problems associated with interpreting practice effects in intact domains versus recovery of impaired domains. From a clinical perspective the preferred outcome is that the stroke patient improves across time, whether or not this improvement can be attributed to practice effects or natural recovery is debatable. The less welcome outcome is where performance does not improve across sessions. If a particular test is insensitive to practice effects in the normal population, no improvement in a stroke victim could well be a sign that cognition in this domain is intact. Alternatively, it could be an impaired function for which there is no recovery. In such instances, the lack of improvement is not diagnostic. Consequently, from a clinical perspective, tasks that show improvement across repetitions in a normal population are preferable to tasks that do not (Darby et al., 2002).

There are a number of ways of determining whether or not there are significant learning (repetition) effects, and these will be explored at greater length in Chapter 5. In the following set of experiments, learning was evaluated by means of ANOVA procedures where improvement was established via a significant main effect for testing session. In the case where there were significant learning effects across sessions, the shape of the learning curve was examined in terms of linear, quadratic and higher order components to establish the trajectory of such learning.

Finally, a number of the cognitive tasks have inbuilt benchmark manipulations (e.g. syllable length in the Anagram task), and all include a dual task format. It is important to know that the benchmark effect is present in the task, and to know how that effect changes

with practice. More importantly, two levels of task difficulty have been implemented in the dual task versions of the task, but as yet these manipulations have not been calibrated in any of the tasks. It is important to establish that people who have recently suffered a stroke are likely to complete the tasks under both levels of difficulty.

The extent to which the four cognitive tasks meet these criteria is explored in the following experiments. In each case, the tasks are explored using a sample of younger, healthy participants. Such a group is likely to give the best indications of the task characteristics. For each task the main interest lies in changes associated with the primary measure.

#### **4.2 Experiment 4.1A**

The Common Associates Test was based on the Compound Remote Associates Test (Mednick, 1962). Each trial in the Common Associates Test consisted of three strong associates of a fourth item. The participant's task was to produce the common associate of the three presented items. Three forms of the task (Form A, Form B, and Form C) were created with each form consisting of two parts; the first consisted of twenty triples and the second consisted of 30 triples. The first set was designed to test event-based monitoring, and the second was used to assess time-based switching performance.

While the three forms were designed to be equivalent there were a number of unknowns. Firstly, it was not known how difficult producing the associates would actually be and if the three forms were actually equivalent in difficulty. Secondly, it was not known how embedding an event detection component within the task would influence production of the associates. Thirdly, it was not known how repeated exposure to the tests would influence performance. The following experiments attempted to address this issue. In Experiment 4.1A the trials were presented in single-task format only. That is, participants only had to generate the common associate. In Experiment 4.1B, the dual task version of the test was examined. In

both instances the participants received all three versions of Common Associates Test and did both Test1 and Test 2 of each Form.

### **4.3 Method**

#### **4.3.1 Ethics**

Approval for this body of research was sought and approved by the Australian Catholic University's Ethics Committee. All research reported in this chapter was collected in accordance with the conditions established under the approval number Q 2011 73.

#### **4.3.2 Participants**

The participants in this study were 52 volunteers (37 female and 15 males) whose ages ranged from 18 to 61 ( $M = 24.9$  years,  $SD = 12.6$ ).

#### **4.3.3 Materials**

The materials were Form A, Form B and Form C of the Common Associates test as described in Chapter 3. For each form, the 20 triples made up Test 1 and 30 triples made up Test-2. In all tests some of the stimulus words were presented in upper case, some were underlined, and some were presented in bold font.

#### **4.3.4 Procedure**

The participants were tested in a group format. Participants were first instructed about the nature of the stimuli, how the three words on each trial were all related to a fourth word, and that their task was to identify the common associate. They were told that there was no correct answer and that they were to write down the first word that came to mind. Participants were told to ignore the fact that some of words would be presented in upper case, bold, italics or underlined.

Although there are effectively six different versions of the Common Associates Test, the order of presentation was not counterbalanced as would normally be expected. The Common Associates Test was designed to be part of a battery of tests and it is standard

clinical practice to maintain a consistent order of testing across participants. Consequently, the tests were presented in the fixed format that corresponded to the planned testing sessions.

Test 1 of Form A was presented first to participants. The test (and all subsequent tests) was presented via Microsoft Powerpoint software with the triples being presented on a large screen such that all triples could be easily read by all participants in the room. Each triple was presented for 6 seconds, and in that time participants had to generate the common associate and to write it down on a prepared answer sheet.

Following the presentation of Test 1 of Form A, participants were given Test 2 of Form A with the same set of instructions relating to the structure of the task and how to respond. Following completion of that test, they were administered the two tests of Form B and the two tests of Form C, in that order.

Three types of responses were possible for each triple. A person could produce the target word. They could produce another word, an intrusion, or they could fail to produce any word, an omission. Omissions accounted for less than 10% of all responses. Target production served as the dependent variable in subsequent analyses.

## **4.4 Results**

### **4.4.1 Response Production**

Performance on the task is presented in Figure 4.1. As far as target recall goes it was clear that in spite of all forms and tests being created to have the same degree of associative strength, when it comes to actual target production, not all tests were equally difficult. It seems that Form C was easier than the others. Moreover, the two tests on each form were quite variable.

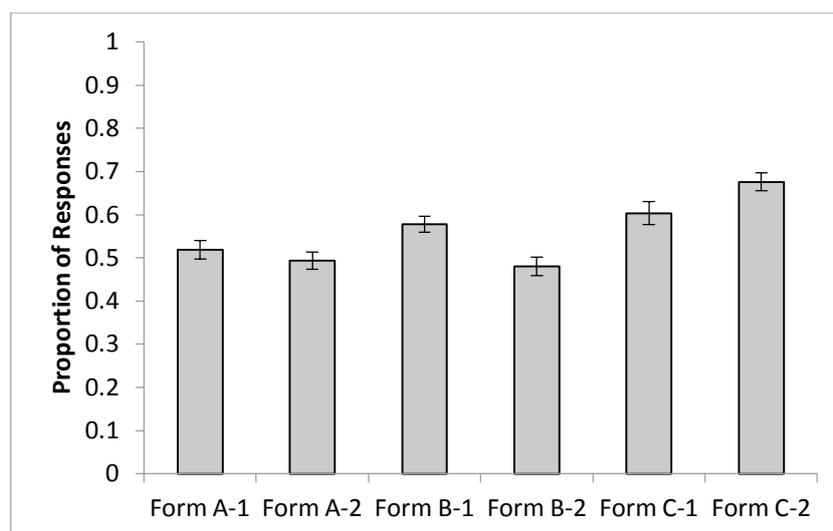


Figure 4.1. Target production on the three forms of the Common Associates Test. Error bars present the standard error of the mean.

The data for target recall were analysed by means of two 3 x 2 repeated measures ANOVA with test form and test number as the respective variables. There was a significant difference between the three forms of the test,  $F(2,102) = 42.14, p < .001, \eta_p^2 = .54$ . Target production was significantly greater on Form C than either Form A,  $F(1,51) = 54.01, p < .001, \eta_p^2 = .51$ , or Form B,  $F(1,51) = 63.24, p < .001, \eta_p^2 = .55$ . Form A and Form B did not significantly differ from each other,  $F(1,51) = 2.59, p = .11, \eta_p^2 = .05$ . There was no significant difference between performance on Test 1 and Test 2,  $F(1,51) = 2.54, p = .12, \eta_p^2 = .04$ , but the interaction between Form and Test was significant,  $F(2,102) = 22.75, p < .001, \eta_p^2 = .31$ . There was no significant difference between Test 1 and Test 2 in Form A,  $t(51) = 1.25, p = .21$ , performance on Test 1 was significantly better than Test 2 in Form B,  $t(51) = 5.83, p < .001$ , and on Form C performance was better on Test 2 than Test 1,  $t(51) = 4.18, p < .001$ .

#### 4.4.2 Reliability

The correlations involving target production between the various tests are summarised in Table 4.1. All correlations were significant at a .05 level.

Table 4.1

*Pearson's Correlation Coefficients for Target Production on Common Associates Test*

Test version	A1	A2	B1	B2	C1
A2	0.53				
B1	0.47	0.57			
B2	0.58	0.67	0.66		
C1	0.39	0.54	0.61	0.71	
C2	0.39	0.64	0.61	0.70	0.76

The intraclass correlations examining the relationship between successive testing sessions are presented in Table 4.2. As is evident in the table, reliability of the different forms of the test was very weak, with the reliability reaching an acceptable level only at the last session.

Table 4.2

*Intraclass Correlations [95% Confidence Intervals] on Target Production of the Common Associates Test*

Measure	A1 & A2	A2 & B1	B1 & B2	B2 & C1	C1 & C2
Target	.53 [.30, .70]	.49 [.26, .67]	.64 [.45, .77]	.69 [.52, .81]	.73 [.56, .83]

## 4.5 Discussion

Three clear findings emerged from the current experiment. The first was that levels of target recall were neither on floor nor on ceiling. The second finding was that the forms were not equivalent in terms of target production. Moreover, the level of target production was not consistently the same for Test 1 and Test 2 of each form. Thirdly, the estimates of alternate-form reliability were quite low, in most instances below the accepted cut-off of .70. The fact that there were alternate form differences and the reliability estimates were so low casts serious doubt on the utility of this measure.

## 4.6 Experiment 4.1B

Experiment 4.1A examined performance on production of the common associate in that absence of any secondary task. In Experiment 4.1B performance was tested when both event-based and time-based monitoring task were included.

## 4.7 Method

### 4.7.1 Participants

Thirty two first-year psychology students (25 females and 7 males, aged between 18 and 58, [ $M = 25.8$ ,  $SD = 11.6$ ]) participated in this experiment for partial course credit.

### 4.7.2 Materials

The materials were identical to those used in Experiment 4.1A.

### 4.7.3 Procedure

As was the case in Experiment 4.1A, the data were collected in a group setting approximately three months after Experiment 4.1A. The procedure was identical to that employed in Experiment 4.1A, except that participants were told about the dual task nature of the experiment. They were told that for Test 1 they were to produce the common associate but in addition they were to monitor for the presence of words presented in capital letters.

The 20 triples were presented in the same way as the previous experiment and participants again had 6 seconds per triple in which to make a response on a prepared answer sheet. The answer sheet had 20 numbered spaces in which to write down the response. Adjacent to each response space was a box containing the letter C. Participants were instructed that on any trial where one of the words was presented in capital letters, they were first to write down the common associate and then circle the C in the box next to where they had written their responses.

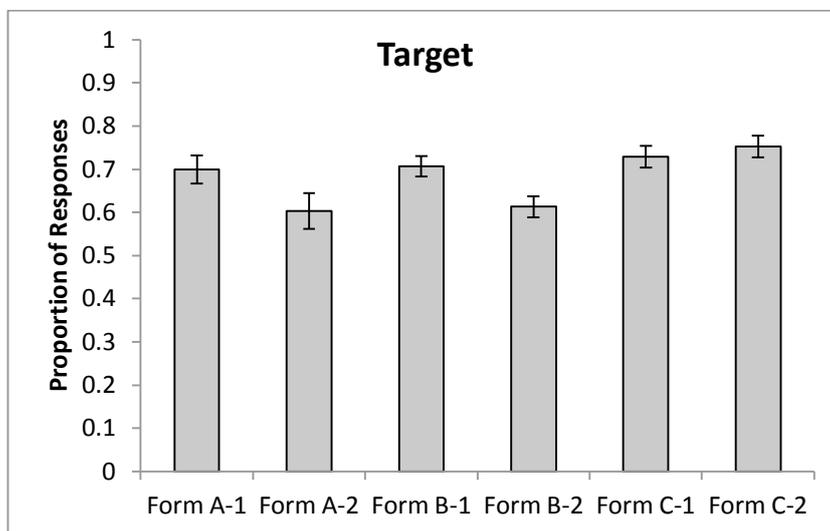
The instructions for Test 2 indicated that they were to search for words in capital letters and for words that were underlined. They were also told that they had to monitor a digital clock that was visible to them, and change their response every thirty seconds. They were to start looking for words that were in capitals, but after 30 seconds to switch to looking for words that were underlined. After a further 30 seconds they were to switch back to looking for words in capitals, and so on. The instructions stressed that on each triple they had to first produce the common associate. From there they had to circle either a C in a box on the response sheet when a word in capital letters occurred, or a U in a box on the response sheet when an underlined word had been presented. The prepared response sheet consisted of 30 numbered places for responses to be written down, together with two adjacent boxes, one containing a C and one containing a U.

## **4.8 Results and Discussion**

### **4.8.1 Production of the Associate**

Performance regarding the production of the associate is presented in Figure 4.2. The data for production of the associate were analysed by means of a 3 x 2 repeated measures ANOVA with Test Form and Test number as the respective variables. For target production there was a significant difference between the three forms of the test,  $F(2,60) = 9.35, p < .001, \eta_p^2 = .24$ . Target production was significantly greater on Form C than either Form A,  $F$

(1,30) = 13.72,  $p = .001$ ,  $\eta_p^2 = .40$ , and Form B,  $F(1,30) = 19.94$ ,  $p < .001$ ,  $\eta_p^2 = .40$ . Form A and Form B did not significantly differ from each other,  $F(1,30) = .10$ ,  $p = .75$ ,  $\eta_p^2 = .01$ . More targets were produced on Test 1 than on Test 2,  $F(1,30) = 13.36$ ,  $p = .001$ ,  $\eta_p^2 = .29$ , and the interaction between Form and Test was significant,  $F(2,60) = 7.07$ ,  $p = .002$ ,  $\eta_p^2 = .19$ . Significantly more associates were produced on Test 1 than Test 2 on Form A,  $t(30) = 2.87$ ,  $p = .007$ , and on Form B,  $t(30) = 3.64$ ,  $p = .001$ , however there was no significant difference on Form C,  $t(30) = 1.38$ ,  $p = .17$ .



*Figure 4.2.* Performance on Associate Production on the Three Forms of the Common Associates Test. Error bars present the standard error of the mean.

Several of the features that were present in Experiment 4.1A were replicated in Experiment 4.1B. Again, production of the associates did not appear to be on ceiling or on floor. Form C of the task appeared to be easier than either Form A or Form B. While the differences between Test 1 and Test 2 were again quite varied, the results were not the same as those in Experiment 4.1A. In Experiment 4.1A there were no differences on Forms A and B and better performance on Test 2 on Form C. In the current experiment performance was better on Test 1 in Forms A and B and there was no difference on Form C. This difference

between Tests 1 and Test 2 reflects different levels of difficulty in the secondary task. In Test 1 target production was conducted against a background of detecting a word in capital letters. In Test 2 participants had to detect two different events and also monitor a clock at the same time.

#### 4.8.2 Single Event Monitoring (Test 1)

Performance on the single event monitoring component of the test is presented in Figure 4.3.

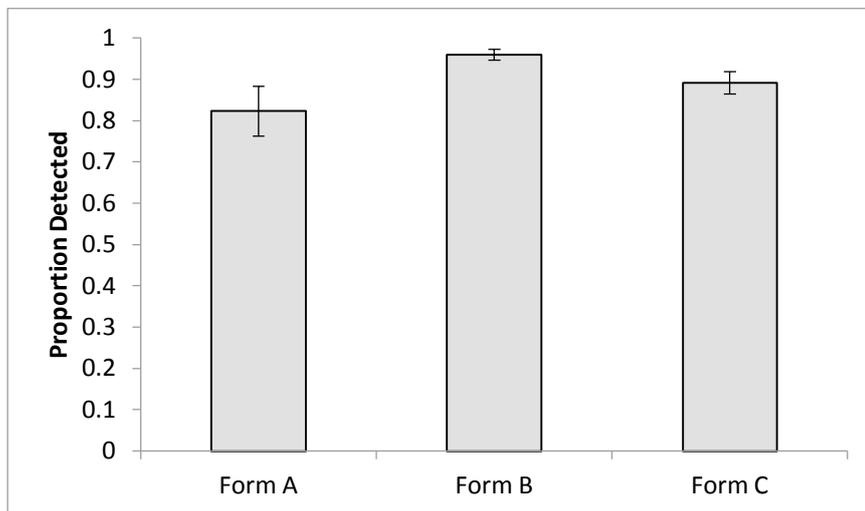


Figure 4.3. Detection of the capitalised word across the three forms of the Common Associates Test. Error bars present the standard error of the mean.

Analysis of the event identification involved a one-way repeated measures ANOVA with the form of the test as the independent variable. As is apparent in Figure 4.3, detection of the event differed significantly across forms of the test,  $F(2,60) = 4.19, p = .02, \eta_p^2 = .12$ . Identification of the event was worse on Form A than on Form B,  $F(1,30) = 5.92, p = .02, \eta_p^2 = .16$ , and better on Form B than on Form C,  $F(1,30) = 10.10, p = .003, \eta_p^2 = .25$ . Form A and Form C did not significantly differ from each other,  $F(1,30) = 1.52, p = .23, \eta_p^2 = .05$ .

The results suggest that participants experienced some difficulty when they first encountered the event detection task, although performance was quite good. By the second time they had done the task, performance was near ceiling. By the third time, errors were made again. The results indicate that the event-based monitoring task did not present any marked difficulties for participants.

#### **4.8.3 Dual-Event Time-Switching (Test 2)**

The time-switching dual-event monitoring task occurred on Test 2 of the three forms. There are three measures of performance: identification of the word in capital letters, which was the response they were required to monitor at the start of the test; identification of underlined words; and the number of switches that participants made between the two identification tasks. These data are presented in Figure 4.4.

Participants' ability to detect the items in capitals and underlined was analysed by a 3 x 2 repeated measures ANOVA with Form and type of detection as the factors. There was no significant difference in detection between the three forms of the test,  $F(2,60) = 1.49, p = .25, \eta_p^2 = .05$ . Participants were better at detecting words in capital letters than underlined words,  $F(1,30) = 5.40, p = .027, \eta_p^2 = .15$ , although this main effect was moderated by a significant interaction,  $F(2,60) = 6.51, p = .003, \eta_p^2 = .18$ . As is evident in Figure 4.4, there was no significant difference in detecting capitalised and underlined words on Form A,  $t(30) = .49, p = .62$ , and in Form B,  $t(30) = .77, p = .45$ , however there was a significant difference on Form C,  $t(30) = 5.02, p = <.001$ .

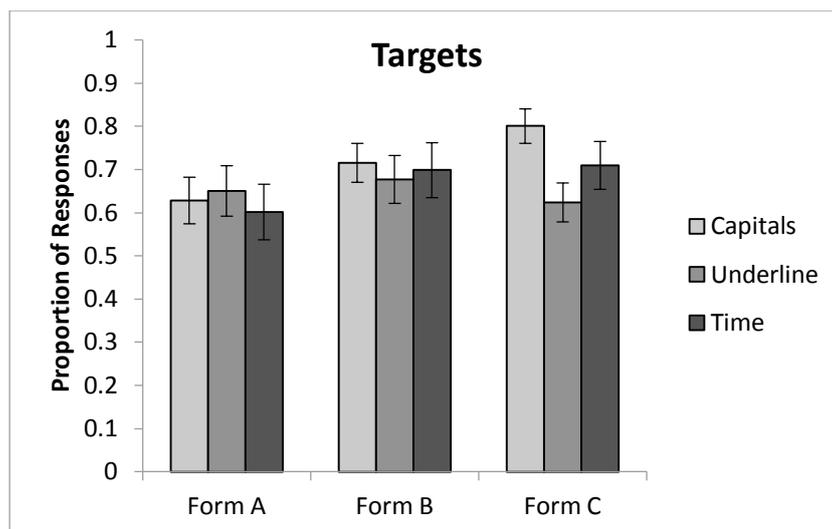


Figure 4.4. Detection of events and time change across the three forms of the Common Associates Test. Error bars present the standard error of the mean.

The participants' ability to monitor and change responses every 30 seconds was analysed with a one-way repeated measure ANOVA. There was no significant difference in this ability across the three forms,  $F(2,60) = 1.89, p .16, \eta_p^2 = .06$ .

Comparison of Figures 4.3 and 4.4 show that in the single-event detection condition (Test 1) participants were able to report the associate and do the secondary task. When the participants were required to monitor for two different events and change which event needed to be monitored every 30 seconds, performance on the monitoring tasks decreased dramatically. In the single-task condition, average detection rates were in the 90% region. In the dual-event plus switch condition, correct event detection and accuracy in making the appropriate switch decreased to the 70% level

#### 4.8.4 Reliability

The correlations in target production among the six tests are presented in Table 4.3. All correlations were significant at the .05 level.

Table 4.3

*Pearson's Correlation Coefficients for Target Production on Common Associates Test*

Tes	A1	A2	B1	B2	C1
A2	0.61				
B1	0.50	0.32			
B2	0.56	0.60	0.43		
C1	0.55	0.60	0.43	0.63	
C2	0.64	0.56	0.44	0.61	0.77

Intraclass correlations for successive alternate forms are presented in Table 4.4. The clear outcome of these analyses was that the test had unacceptable reliability.

Table 4.4

*Intraclass Correlations [95% Confidence Intervals] for Target Production on the Common Associates Test*

Measure	A1 & A2	A2 & B1	B1 & B2	B2 & C1	C1 & C2
Target	.59 [.31, .78]	.27 [-.08, .57]	.42 [.09, .67]	.62 [.35, .80]	.76 [.57, .88]

#### 4.9 General Discussion of Experiment 4.1

The first experiment was informative in relation to a number of aspects of the newly developed Common Associates Test. The test produced acceptable levels of target production (56% and 68%) with performance being above floor and below ceiling. The production rates were higher than the average pre-existing associative strength of the three items to the target item. In developing the Common Associates Test, the three forms were matched for pre-

existing strength between the three associates and the target item with the expectation that actual performance across the three tasks would be equivalent. This expectation was not realised. In both Experiment 4.1A and 4.1B, performance on Form C was better than performance on Forms A and B.

Adding a single-event detection component to the Common Associates Test did not appear to have any great impact on production of the associate, and detecting the event was excellent after the first time doing the task. On Test 2, adding two detection tasks and a time component to the detection task made target production, event detection and time monitoring more difficult. In fact, feedback from participants indicated that this version of the task was very demanding, and as such it is questionable as to the appropriateness of manipulating task difficulty in this way for patients with stroke.

The last and probably most critical feature of the results was that reliability estimates were unacceptably low. In both parts of the experiment, the .70 criterion was only reached between the second last and last sessions. This fact combined with the presence of systematic differences between the alternate forms and the level of difficulty associated with the dual-event time switching version of the task means that the task failed on a number of the criteria for inclusion in the battery. Thus, while the task was designed to measure one aspect of cognition, it failed to do so in an acceptable manner and therefore was deleted from the battery.

#### **4.10 Experiment 4.2**

Experiment 4.2 explored performance on the three remaining cognitive tasks, namely, the Snap, Immediate Serial Recall and Anagram tasks. These tests were administered in their single-task format to a group of young, healthy participants on six occasions over a 2-week period with a number of aims in mind. The first was to explore the change in performance across testing sessions. The primary measure of each test was analysed in terms of the degree

of linear and non-linear changes in performance across the six testing sessions. Secondary analyses involved replication of benchmark findings where appropriate. Thus, in the Anagram task, it was expected that solution times for 7-letter words should be faster than for 6-letter letter words, which in turn should be faster than for 5-letter words (Muncer & Knight, 2011). Similarly, lists of associatively related items should be better recalled than lists containing unrelated items in immediate serial recall (Tehan, 2010; Tse, 2009). The third aim was to evaluate the psychometric properties of the tests. Internal consistency measures (Chronbach's alpha) were calculated where possible (anagram) and test-retest measures were also reported in terms of bivariate correlations among the 6 test sessions and by intraclass correlation coefficients of successive testing sessions.

## **4.11 Method**

### **4.11.1 Participants**

The participants in the study were 24 volunteers (12 male and 12 female) whose ages ranged from 18 to 48 ( $M = 31.63$ ,  $SD = 10.17$ ).

### **4.11.2 Materials**

The tests used in this experiment were Snap, Anagram and Immediate Serial Recall tasks as outlined in Chapter 3.

### **4.11.3 Procedure**

All tests were administered in their single-task configuration. That is, there were no dual task conditions. The three tasks were administered in a single session of each testing occasion. All participants were tested on the battery six times across a 2-week period. They were tested three times in the first week and three times in the second week.

## 4.12 Results

### 4.12.1 Snap Task

**4.12.1.1 Repetition and benchmarks.** The completion times on the Snap task are presented in Figure 4.5.

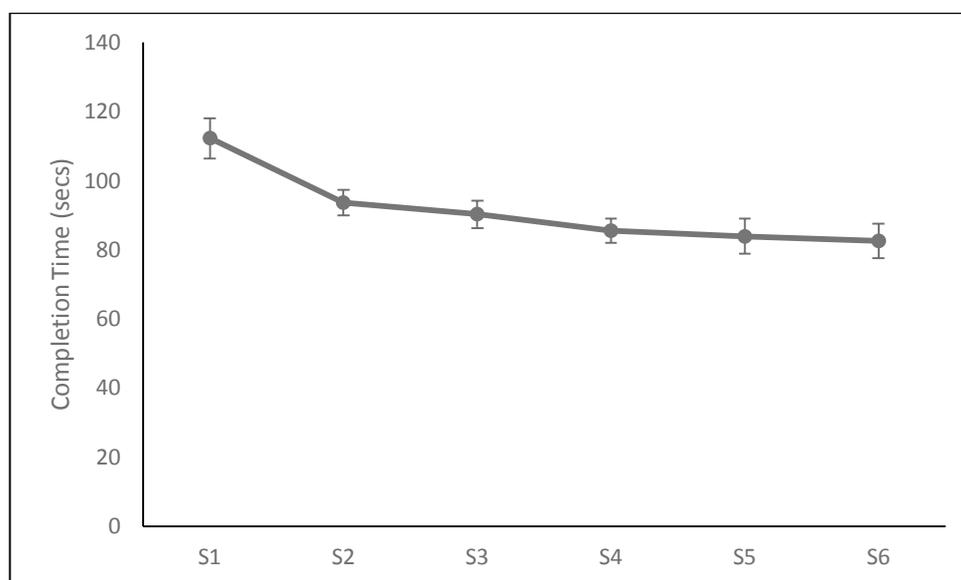


Figure 4.5. Completion time on the Snap task as a function of testing session. Error bars represent the standard error of the mean.

The data were analysed by means of a one-way repeated measures ANOVA. There was a significant improvement in completion times across the testing sessions,  $F(5,115) = 14.79, p < .001, \eta_p^2 = .39$ . The improvement in performance had a strong linear component,  $F(1,23) = 28.54, p < .001, \eta_p^2 = .55$ , and a significant quadratic component,  $F(1,23) = 8.10, p = .009, \eta_p^2 = .26$ , due to the marked increase from the first to second testing sessions.

**4.12.1.2 Reliability.** The simple correlations between the testing sessions are presented in Table 4.5, and reliability estimates based on intraclass correlations are presented in Table 4.6. All bivariate correlations in this analysis, and all subsequent ones, were significant at the .05 level unless otherwise indicated. Intraclass correlations were significant unless otherwise indicated.

Table 4.5

*Pearson's Correlation Coefficients for Six Testing Sessions on the Snap Task*

Session	S1	S2	S3	S4	S5
S2	0.44				
S3	0.40 <sup>NS</sup>	0.86			
S4	0.45	0.79	0.84		
S5	0.52	0.73	0.67	0.82	
S6	0.44	0.61	0.63	0.73	0.92

Note: NS = non-significant at .05 level.

Table 4.6

*Intraclass Correlations [95% Confidence Intervals] on the Snap Task*

Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Snap	.40 [.00, .68]	.86 [.70, .93]	.83 [.65, .92]	.76 [.52, .89]	.91 [.81, .96]

It is apparent in all measures that performance was more variable on the first test session than on the remaining sessions. By the second test performance had stabilised, and remained stable across the remaining five sessions.

#### 4.12.2. Anagram Task

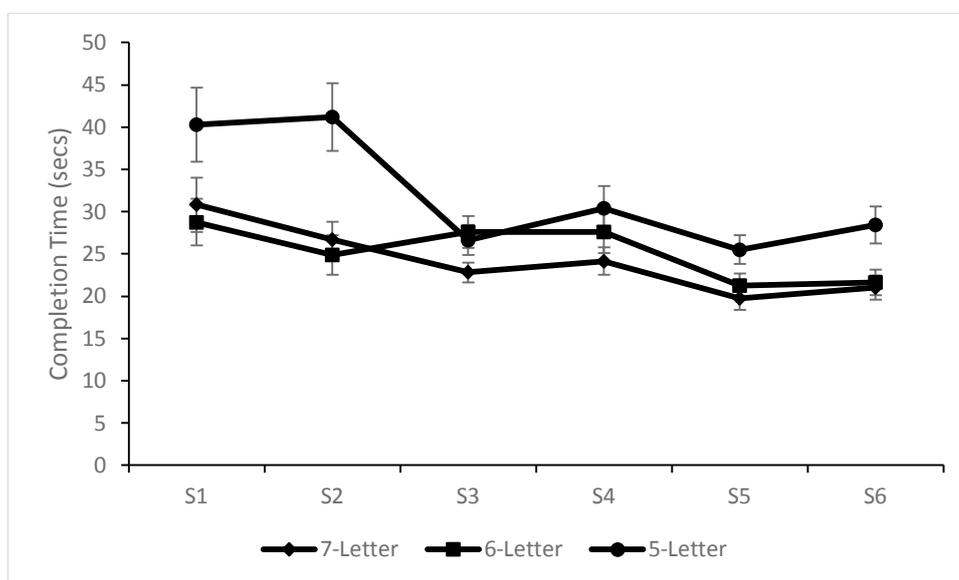


Figure 4.6. Completion time on the Anagram task as a function of testing session and letter length. Error bars represent the standard error of the mean.

**4.12.2.1 Repetition and benchmarks.** The Anagram task results are summarised in Figure 4.6. These data were subjected to a 6 session by 3 letter length repeated measures ANOVA. There were significant effects of letter length,  $F(2,46) = 18.21, p < .001, \eta_p^2 = .44$ . There was no difference between 7-letter completion times and 6-letter completion times,  $F(1,23) = 3.75, p = .065, \eta_p^2 = .06$ , but 7-letter anagrams were completed significantly faster than 5-letter anagrams,  $F(1,23) = 19.65, p < .001, \eta_p^2 = .46$ .

There were also significant differences across testing session,  $F(6,115) = 12.05, p < .001, \eta_p^2 = .34$ . There was a strong linear decrease in completion times across sessions,  $F(1,23) = 21.84, p < .001, \eta_p^2 = .49$ , and there was a significant 5<sup>th</sup> order component,  $F(1,23) = 12.54, p = .002, \eta_p^2 = .35$ .

The interaction between length and session was also significant  $F(10,230) = 7.24, p < .001, \eta_p^2 = .24$ . As is evident in Figure 4.6 length differences were more apparent in the early testing sessions than in the latter ones. More importantly however, there were significant linear effects for 7-letter,  $F(1,23) = 18.27, p < .001, \eta_p^2 = .44$ , 6-letter,  $F(1,23) = 7.68, p = .011, \eta_p^2 = .25$ , and 5-letter anagrams,  $F(1,23) = 14.79, p < .001, \eta_p^2 = .39$ .

The finding that 5-letter anagrams were significantly slower to complete than 6-letter and 7-letter anagrams confirmed the manipulation of task difficulty, and was therefore consistent with benchmark expectations. More importantly, as participants had more exposure to this task they became faster, as reflected in the strong linear component in improvement. While the 5<sup>th</sup> order component was statistically significant it is not easy to interpret, and it explains less variance than the linear component.

**4.12.2.2 Reliability.** Simple and intraclass correlations between the testing sessions are presented in Tables 4.7 and 4.8 respectively where performance has been collapsed across letter length to form a completion time total score. Internal consistency measures for each session, based on completion times for 7-letter words, 6-letter words and 5-letter words, are presented in the first row of Table 4.7. By all measures the test has acceptable reliability.

Table 4.7

*Pearson's Correlation Coefficients for Six Testing Sessions on the Anagram Task*

Session	S1	S2	S3	S4	S5	S6
Alpha	.90	.70	.83	.82	.84	.79
S2	0.79					
S3	0.84	0.84				
S4	0.68	0.73	0.76			
S5	0.75	0.74	0.67	0.73		

S6      0.80      0.75      0.75      0.79      0.90

Note: All correlations significant at .05 level

Table 4.8

*Intraclass Correlations [95% Confidence Intervals] on the Anagram Task*

Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Anagram	.76 [.52, .88]	.72 [.45, .86]	.72 [.45, .86]	.69 [.37, .54]	.89 [.76, .95]

### 4.12.3 Immediate Serial Recall Task

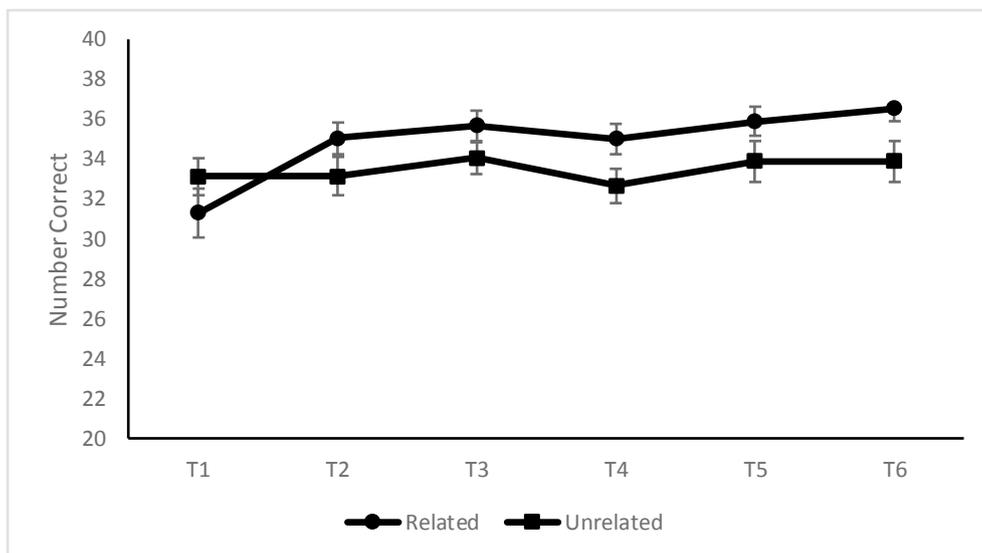


Figure 4.7. Number of correctly recalled items on the Immediate Serial Recall task as a function of associative relatedness.

**4.12.3.1 Repetition and benchmarks** The Immediate Serial Recall task results are summarised in Figure 4.7. These data were subjected to a 6 session by 2 relatedness repeated

measures ANOVA. There were significant effects of relatedness,  $F(1,24) = 9.58, p = .005, \eta_p^2 = .29$ . There were also significant differences across testing session,  $F(5,115) = 6.63, p < .001, \eta_p^2 = .2$ . The interaction between relatedness and session was also significant  $F(5,115) = 4.19, p = .002, \eta_p^2 = .15$ . As is evident in Figure 4.7, the interaction was primarily due to unexpectedly poor performance on the related lists in Session 1. Inspection of the simple main effects showed that with the unrelated lists, there was no significant effect of testing session,  $F(5,115) = 1.01, p = .41, \eta_p^2 = .04$ . In the related lists, when only Sessions 2 to 6 were considered, there was no significant effect of test session,  $F(4,92) = 1.47, p = .21, \eta_p^2 = .06$ .

**4.12.3.2 Reliability.** Simple and intraclass correlations between the testing sessions are presented in Tables 4.9 and 4.10 respectively for both associatively related lists and unrelated lists. The correlations in Table 4.9 were generally lower than expected. More importantly, the intraclass correlations were all at an unacceptably low level, with the possible exceptions being Sessions 3 to 5 with unrelated lists. Even in these instances the reliability only just meets criterion.

Table 4.9

*Pearson's Correlation Coefficients for Six Testing Sessions on the Related Lists in Immediate Serial Recall.*

Trial Type	S1	S2	S3	S4	S5
<i>Related</i>					
S2	0.24 <sup>NS</sup>				
S3	0.30 <sup>NS</sup>	0.44			
S4	0.55	0.28	0.51		
S5	0.64	0.55	0.62	0.59	
S6	0.44	0.47	0.60	0.51	0.49
<i>Unrelated</i>					
S2	0.65				
S3	0.57	0.49			
S4	0.74	0.59	0.73		
S5	0.76	0.62	0.66	0.78	
S6	0.65	0.70	0.61	0.61	0.64

Note: NS = non-significant at .05

Table 4.10

*Intraclass Correlations [95% Confidence Intervals] on the Immediate Serial Recall Task*

Trial Type	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Related	.21 <sup>NS</sup> [-.19, .56]	.43 [.05, .71]	.51 [.14, .75]	.58 [.24, .79]	.48 [.11, .73]
Unrelated	.64 [.33, .83]	.48 [.10, .73]	.72 [.45, .87]	.76 [.53, .89]	.63 [.32, .82]

Note: NS = non-significant at .05

#### 4.13 Discussion

This experiment explored the characteristics of the remaining three tasks used in the cognitive component of the study. The tasks were presented in their single task configuration to examine how the characteristics of the tasks changed with repeated exposure, prior to implementing a secondary task in the dual task versions. The three primary considerations were the extent to which performance improved with practice, the degree to which benchmark effects would be apparent and how those effects changed with practice, and whether the psychometric properties of the test were adequate.

In the case of the Snap task, learning effects were present in the data in that there was a strong linear decrease in completion time across the six sessions. The improvement was most pronounced from Session 1 to Session 2. The psychometric properties of the task were also excellent with the exception of the test-retest effect for Sessions 1 and 2. This test meets the established criteria.

The Anagram task also met the pre-determined criteria. Completion time improved with practice on the task, and this was true for anagrams of all lengths. The expectation that

solution times for 7-letter anagrams would be faster than for 6-letter anagrams and that 6-letter times would be faster than for 5-letter anagram times was partially supported (Muncer & Knight, 2011). There was not a significant difference between 7- and 6-letter anagrams, but both were solved significantly faster than 5-letter anagrams. More importantly, with repeated practice the effect of anagram length was attenuated. The improvement in completion time for five-letter anagrams was more pronounced than for the longer anagrams. Finally, the psychometric properties of the task were acceptable.

The Immediate Serial Recall task on the other hand, did not meet two of the criteria. Performance on the associatively related items was superior to that of the unrelated lists, replicating the Tse (2009) and Tehan (2010) research. However, there was no increase in performance across the sessions in the unrelated conditions, and none in the related conditions if the aberrant first session is ignored. In addition, the psychometric properties of the tasks were not adequate. The failure on both the learning and psychometric criteria indicates that the test should be omitted from the battery.

#### **4.14 General Discussion**

Experiments 4.1 and 4.2 dealt with the characteristics of the four cognitive tasks designed to be used in the study. Three criteria were used to evaluate the tasks, namely the reliability of the measures, the degree to which repetition (practice) of the task produced improved levels of performance, and the degree to which benchmark phenomena that were included in the design of the tasks were realised in the data. A fourth consideration, dual task difficulty was also examined in Experiment 4.1.

Experiment 4.1 explored the newly developed Common Associates Test in some depth. While it was clear that participants could do the task, in that there were neither ceiling nor floor effects on production of the associates, two of the criteria were not met. Firstly, there were significant differences between the three different forms of the test, although all

three had been designed such that associative strength between the three test items and the target response were equivalent based on norms of associative strength (Nelson, McEvoy, & Schreiber, 1998). The non-equivalence of these forms compromises the ability to detect learning effects across repeated testing sessions where alternate forms would be in use. Not only were there differences in mean levels of performance, but the alternate form reliability was also unacceptably low. While mean differences might change between tests, the relative order of the participants could remain the same resulting in high reliability between the alternate forms. This did not occur in the current data; the low reliability indicates that one person could do well on one form of the test but do poorly on the alternate forms. On the basis of failure to meet these two critical variables, the decision was made to delete this task from the battery.

The Common Associates Test task, in spite of its deficiencies, was useful in calibrating task difficulty in the dual task conditions. When a single detection task was employed, participants were able to produce the common associate and complete the detection task at the 90% level. However, when time-based task switching was implemented, performance on the detection task dropped considerably to between 65% and 70%, in most instances. While this clearly establishes the effectiveness of the task difficulty manipulation, all the participants indicated that monitoring the list items and the clock, and remembering to switch from underlined to uppercase was cognitively very demanding. The reported difficulty with the task, together with the level of performance among young healthy participants raised the possibility that with patients with stroke in the early weeks of recovery, performance in this condition could be on floor. Thus, the data supported the continued use of the single-event detection task, but indicated that some modification of the two-event task switching condition was required. The modification employed in the final test battery was to have participants search for two events, without the need to switch from searching for one type of

event to the other. This modification is described in the dual task components of the Snap, Immediate Serial Recall, and Anagram tasks in Chapter 3.

Experiment 4.2 evaluated the three remaining cognitive tasks using the same criteria. In the case of the Anagram and Snap tasks, reliability indices were all within acceptable limits, and performance on the tasks improved with repeated practice. In the case of the Anagram task, bench mark effects were observed in the early test sessions, in that solution times for 7- and 6-letter words were substantially faster than for 5-letter words. With practice this difference in completion times was attenuated. Since both tasks met the established criteria, they were used in the final battery.

The Immediate Serial Recall task also failed to meet expectations on two of the criteria. The first problem involved the lack of improvement across testing sessions. The lack of repetition effects in a normal sample means that lack of improvement with practice in a stroke patient is not diagnostic of failures in the recovery process. The second issue was again the lack of test-retest reliability. The intraclass correlations were unacceptably low. While the benchmark finding of better recall for lists containing associatively related items was superior to that of lists containing unrelated items, this positive aspect of task performance did not outweigh the reliability and practice issues associated with the task. Consequently, the failure to meet two key criteria meant that the Immediate Serial Recall task had to be eliminated from the cognitive component of the test.

The failure of the Common Associates Test and Immediate Serial Recall tasks to meet key requirements has two implications. Firstly, the number of cognitive tests in the battery is reduced to two tests only. Secondly, the planned use of associative relatedness as a mechanism for augmenting the recovery process has to be abandoned at this stage.

#### 4.15 Chapter Summary

The chapter reported two studies that examined the characteristics of the four cognitive tasks. Criteria for using the tasks were determined with an emphasis on psychometric reliability, the presence of practice effects, equivalence of alternate forms where applicable, and the presence of benchmark effects where applicable. The Common Associates Test and the Immediate Serial Recall tasks failed to meet these criteria and were deleted from the planned battery. Secondly, in the dual task conditions, the use of a time-based task switching manipulation of task difficulty was deemed to be too difficult for a stroke population.

At the conclusion of this chapter, the test battery consisted of Snap, Anagram, Digit Span, Stroop, Rey Tangled Lines, Phonemic Fluency and Semantic Fluency tasks. Using a single-event detection task was deemed appropriate. A two-event detection task was adopted as a more difficult, but achievable, manipulation of task difficulty in the dual task versions of each test. The next chapter builds on the current results by again presenting the test battery in both single and dual task formats to a group of healthy, older participants who provide a source of normative data. Again, with this group it is important to confirm the psychometric acceptability and benchmark effects that were present in the initial examination of tasks in the battery where the participants were younger healthy adults.

## Chapter 5: Normative Data for the Full Battery

### 5.1 Chapter Overview

The previous chapter explored the four cognitive tests that were planned for use in the test battery. Those tasks were evaluated using young healthy participants who were not necessarily representative of patients with stroke in terms of age and cognitive abilities. It is well documented that there are age differences in cognitive ability, such that there is a general decline across a range of “fluid” cognitive tasks that appears to accelerate when people are in their 70s (Anstey & Low, 2004). Thus, while the young participants were appropriate for evaluating the psychometric and learning aspects of the cognitive tasks, they are not appropriate for cognitive changes following stroke given that most patients with stroke are over 70 years old (Australian Institute of Health and Welfare, 2013).

From a neuropsychological assessment perspective, impairment is typically evaluated by comparing the performance of the individual to large normative samples. Depending on the construct measured, raw scores are converted in to scaled scores (Mean = 10, SD = 3), T scores (Mean = 50, SD = 10), or standard scores (Mean = 100, SD = 15), and for each type of score percentile ranks can be computed. These scores are often adjusted for demographic variables like gender, age and education levels. The process is adopted so that performance on different tests can be compared using a common metric. In the case of the current test battery, norms are available for some of the tests but not others. There are published norms for forward and backward digit span (Wechsler, 2008), the Stroop Colour Word Task (Ivnik et al., 1996), phonemic fluency total scores (Tombaugh et al., 1999), total scores on Animals trial of the semantic fluency task (Tombaugh et al., 1999), and the Rey Tangled Lines Test (Senior et al., 1998). There are no published norms for the Snap and Anagram tasks, nor for the combined three trials in the semantic fluency task, nor for performance on first and later response measures in both forms of the verbal fluency task.

In response to the absence of norms on all tasks, the process adopted in many of the studies exploring prevalence rates described in Chapter 2 are employed here. In those studies a control group of non-stroke, healthy participants took the same test battery as the patients with stroke, and consequently this group served as a de facto normative sample. Australian Bureau of Statistics data indicates that 87% of strokes occur in those who are 55 years and older (Australian Institute of Health and Welfare, 2013). Consequently normative data were collected on a group of participants in this age bracket who were living independently in the community and reported no major health issues when tested. This chapter describes the outcomes of administering the task battery across six testing sessions, and thus provides the normative data for all tasks and all measures with which the performance of patients with stroke can be evaluated.

The current chapter has two other aims. Firstly, given the clinical issue of evaluating performance at the individual rather than the group level, the issue of how to evaluate individual learning across test repetitions needs to be resolved. Secondly, Chapter 4 dealt with the psychometric properties of the cognitive tests in the battery but did not examine the properties of the remaining tasks in the battery. The psychometric properties of all the tests, therefore, are examined using the older sample with the intent of confirming the appropriateness of this group as a normative sample.

## **5.2 Characteristics of Normative Group**

Decisions concerning the desired characteristics of the control/normative group are reliant on the empirical questions that are being addressed and the design of the experiment being conducted. In the case where groups are being assessed using experimental procedures, it is desirable that an experimental group and control group be matched on as many characteristics as possible, so that the effects of confounding variables are minimised. Consequently, in these designs there are usually strict exclusion criteria for the selection of control participants. In studies exploring the effects of stroke, control groups will usually be

matched for age and gender, and there will be exclusion criteria regarding medical history, medication, alcohol and drug usage, mood, and mental health. However, controlling for all possible covariates can be problematic with more than just a few covariates, as it becomes very difficult to find close matches on all the covariates (Stuart & Rubin, 2007). Where close matching is possible, under some circumstances matching can lead to substantial bias in the estimated treatment effect (Frangakis & Rubin, 2002). As indicated earlier, the inability to control for both impairment profile and comorbidities was a strong determinant of adopting a case study approach in the current thesis.

If the emphasis is on evaluating the performance of individual cases, as is the case in the current thesis, then the role of the large control group is to provide normative data with which performance of a stroke patient can be compared. To be useful, the normative sample needs to reflect the population from which the sample is derived. The normative sample should reflect the variability that exists in the relevant population. From this perspective a minimum number of exclusion criteria should apply. Thus, the current sample consisted of a sample of 55 years and older who were living independently in the community. The only exclusion criterion applied was that participants could not have a current diagnosis of any condition that had cognitive impairment as one of the standard symptoms.

In Chapter 2, the studies that explored the prevalence of stroke all based their decisions concerning impairment on the basis of statistical information, most frequently using a cut-off of 1.96 standard deviations from the mean (Hurford et al., 2013; Sachdev et al., 2004a; Tatemichi et al., 1994). Likewise, those studies that explore recovery also used statistical decision making processes (Desmond et al., 1996; Sachdev et al, 2004b). The effect of embracing variability, rather than controlling for it, has a number of consequences when cognitive impairment and learning decisions are based on statistical decision criteria. Firstly, the greater the variability in performance, the more extreme the deviation from normal must be for performance to be classified as impaired. This means that some participants might

actually have cognitive impairments, but not of a sufficient severity to be detected. However, if an impairment is indicated in the data, there is little chance that such an outcome is a false alarm. Secondly, the ability to detect improvement in performance becomes more difficult when evaluated in terms of statistical cut-offs. Improvement across sessions needs to be substantially greater to be recognised as improvement than would be the case if variability among scores were reduced.

A third consequence in using a normative sample rather than a control group is that in the absence of pre-morbid estimates of cognitive abilities, it is not possible to conclude with 100% certainty that exceeding a specified cut-off represents the outcome of having a stroke. By definition, a small percentage of people will show the same extreme scores without any damage to the brain. Moreover, other non-controlled factors could also result in decrements in performance. In the case where an impairment is detected after a stroke, it is highly likely that the impairment is due to the stroke. There is a chance, however, that there are alternative explanations for such a deficit.

In sum, choosing the type of comparison group with which to compare the performance of patients with stroke involves cost-benefit trade-offs. Using highly matched controls makes interpretation easier, but matching for all possible variables is virtually impossible. The use of a large normative sample, as is typically used in neuropsychological assessment, has its benefits. Increases in variability and in not controlling for possible confounds, however, means that it is harder to detect performance decrements and more difficult to provide an unambiguous explanation for changes in behaviour.

## **5.3 Evaluating Improvement**

### **5.3.1 Slopes and Intercepts**

In Chapter 4 the issue of detecting improved performance across testing sessions was addressed by examining main effects for testing session that emerged from repeated-measures ANOVAs. Specifically, learning was indicated by the extent to which the data

could be described by linear or higher order functions across sessions. In the current experiment this process will be used again to evaluate repetition effects within the normative sample.

However, ANOVA procedures are not applicable in evaluating learning in single cases. Alternative measures of evaluating changes in performance must be adopted. The first measure is derived directly from the ANOVA outcomes. A significant linear component of a main effect indicates that learning can be described in terms of a linear function, with each testing session producing the same increment in performance as on previous sessions. Any linear function can be reduced to two parameters, a slope and an intercept. In the current examples, the slope represents the degree of learning that occurs across sessions with slopes of zero indicating that no learning has occurred.

Using linear estimates of changes in performance can be applied to an individual's learning outcomes by calculating the best-fitting regression lines for each test. Again the slope of the linear function can provide an index of learning. To illustrate this, the data (solid lines) for four fictitious participants are represented in Figure 5.1. Best fitting linear regression lines (dotted lines) have been generated for each data set and are plotted together with the raw data. The regression equation for each line is presented as well as an  $R^2$  value. This latter value indicates the amount of variance shared by the actual data and the best-fit line. High values indicate regular improvement across testing sessions.

As indicated in Figure 5.1, the data and regression lines for S1 are hard to distinguish and this is reflected in an  $R^2$  value of .99. The degree of learning is represented by the slope, which has a value of 5.17. S2 does not show as much improvement across sessions as reflected in a lower slope value of 4.11, but the  $R^2$  value is still quite high at .89. S3 has an equivalent slope to S2 of 4.11, but the degree of learning across sessions is not uniform as indicated by the  $R^2$  value of .50. S4 is showing very little learning at all. Consequently, the degree of learning across a number of sessions can be reduced to a single number, the slope

of the best fitting regression line. Any positive measure of slope indicates that some learning is occurring, whereas a zero slope would suggest that there was no learning going on, and a negative slope would indicate that performance deteriorated across sessions.

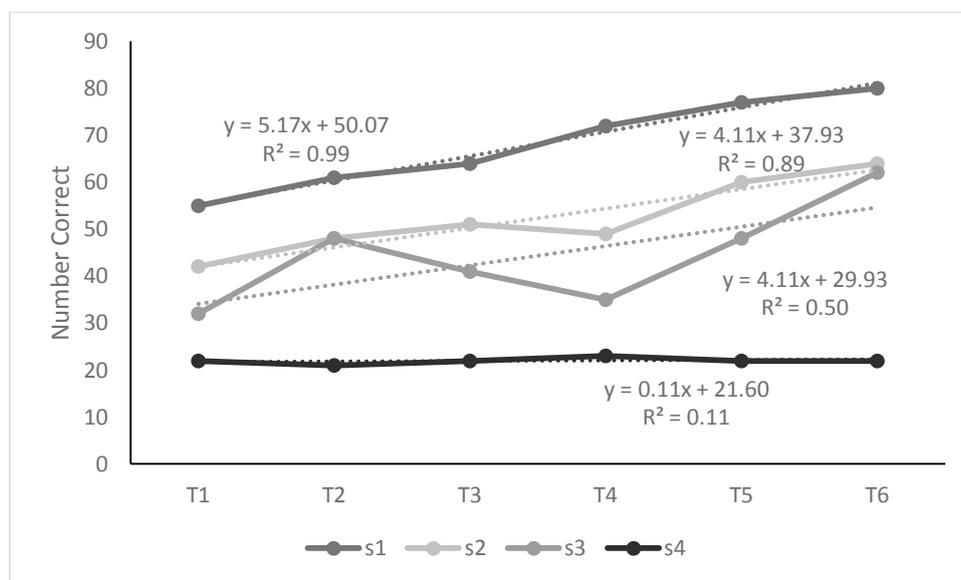


Figure 5.1. Training outcomes (slopes) of four hypothetical participants.

As was the case with absolute levels of performance, one can compare individual estimates of the slope with those obtained by a normative sample by constructing 95% confidence intervals around the average slope of the normative group. Even if a stroke patient shows positive learning outcomes as reflected in positive slope values, the specific value of the slope could still be outside normal range functioning. The effect of learning in the patients with stroke in this study will be assessed using this method of evaluating learning outcomes.

The regression equations not only provide a measure of learning as reflected in the slope, but they also provide an intercept measure as well. The importance of considering the relationship between the slope and the intercept can be seen in Figure 5.2, where performance on a hypothetical speeded task for four hypothetical participants has been plotted. In that Figure, all participants show learning effects in that they are all getting faster across sessions, and all appear to be doing the task as well as possible by the final session. However, the

slopes for the different participants are different suggesting that S4 shows the greatest improvement when compared with the other participants. However, it is also the case that S4 has the greatest room for improvement, because he or she was slowest at the beginning. S1, on the other hand, was fast to start with and could not improve performance all that much on subsequent sessions. Therefore, steep slopes do not necessarily reflect true learning ability because it was not possible for S1 to improve greatly across sessions.

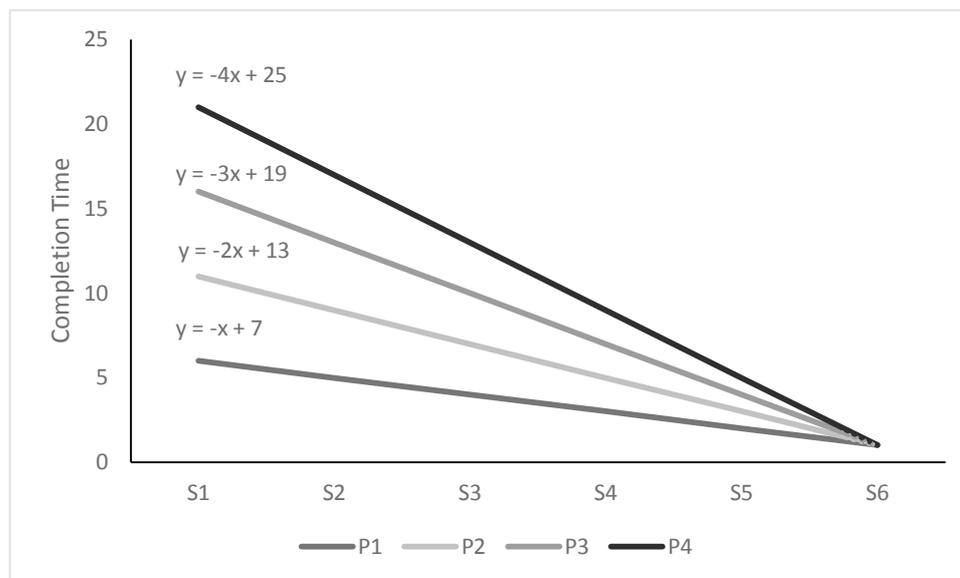


Figure 5.2. Training outcomes (slopes and intercepts) of four hypothetical participants.

In examining Figure 5.2, it is apparent that there is a correlation between slopes and intercepts reflecting the fact that those who were slowest to start with showed the greatest learning. In Figure 5.1 there is no correlation between slopes and intercepts, and as such one can conclude that any changes in performance were not due to initial performance levels. When using slopes to assess learning in any task, an examination of the correlation between slopes and intercepts is necessary to see that learning outcomes are not simply a function of initial performance levels.

### 5.3.2 Temporal stability confidence intervals

The slopes of best-fitting regression lines provide one index of the degree of learning. In the case where learning does occur, however, the slope provides little information

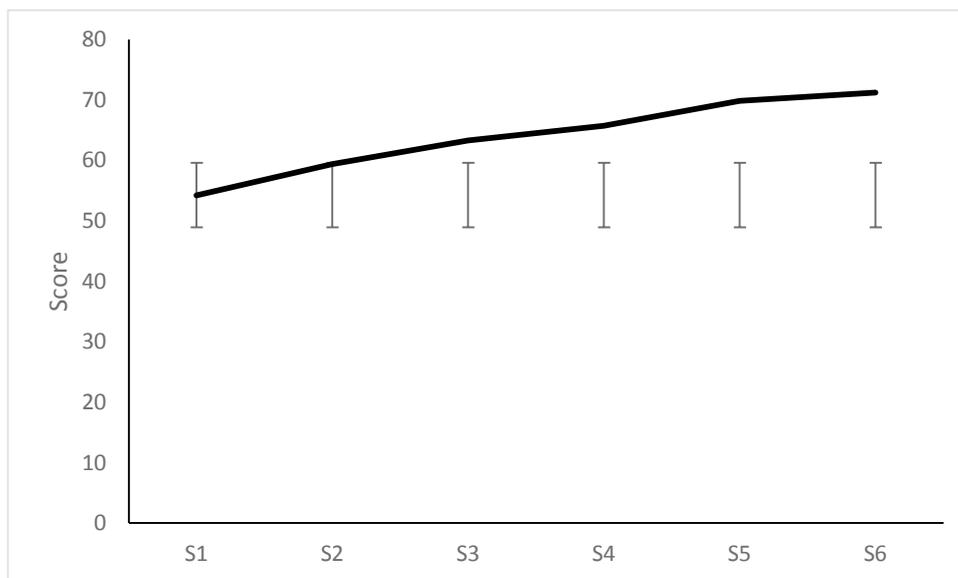
concerning the point at which such learning becomes statistically significant. An alternative method of determining the extent of learning can be derived from examining practice effects when neuropsychological tests are repeated.

The basis for this measure of learning builds on the psychometric properties of the test. As any cognitive test does not have a reliability of 1.0, there is some uncertainty about the true score obtained from the test because of measurement error (and other sources of error). Because of measurement error, the score on a second administration of the test is unlikely to be identical to that obtained on the first test administration. The variability of scores with repeated testing is directly related to reliability. The greater the reliability, the less the variability among repeated administration. In short, those tests that have high reliability are said to exhibit high levels of temporal stability (Strauss et al., 2006).

The degree of error associated with any test can be calculated, as can confidence intervals around the obtained test score. The temporal stability confidence intervals define the range of scores that are likely to occur as a result of measurement error, and these confidence intervals would apply to all subsequent testing sessions. In the case of repeated testing where no learning takes place, the obtained scores should be contained within the confidence interval limits. However, in the case where there is systematic variance due to learning, then the obtained score should at some point move outside the confidence limits.

In applying these principles to an individual case where both the reliability of test and the standard deviation of population (of a large normative sample) is known, temporal stability confidence intervals can be constructed using the standard error of prediction (Lord & Novick, 1968). The confidence intervals are placed around the score for the initial test measure, and the band defined by the confidence intervals on this first test forms the decision making basis for all further tests. For the current battery, confidence bands based on the standard error of prediction could be calculated on scaled scores for forward and backward digit span, the three Stroop measures, total scores for phonemic fluency, the Animal Naming

sub-test of the semantic fluency task, and the Rey Tangled Lines test. This is because reliability and population standard deviations have been previously established. However, for the Anagram, Snap, first and later responses of the phonemic fluency task, and all three measures of the semantic fluency tasks, there are no normative data available on which to calculate the standard error of prediction. Therefore confidence intervals based on the standard error of prediction are not possible.



*Figure 5.3.* Learning effects based on temporal stability confidence intervals.

In the case where reliability and population standard deviations are not known, confidence intervals can still be constructed using the standard error of measurement as a fall-back procedure. Again, confidence intervals are determined for the initial test and are used on subsequent tests to evaluate the presence of learning. Figure 5.3 shows the performance of a hypothetical participant across six testing sessions on a cognitive task. Using the standard error of measurement derived from a normative sample on the same task, 95% confidence intervals have been placed around initial test performance for this participant. The area between the upper and lower limits represents the range of scores that would encompass changes on subsequent tests that were attributable to error. Points outside the confidence band represent systemic changes in performance. The second session scores in Figure 5.3 are on

the boundary of the upper limit, but on subsequent tests the scores are well outside the confidence band and are taken as evidence for the presence of learning.

The data in Figure 5.3 could just as easily represent the means of a sample of participants, in which case the data would imply that by the third testing session, average scores have improved significantly above baseline levels. These outcomes can be directly compared to ANOVA outcomes in which post-hoc analyses comparing initial test performance to performance on subsequent sessions would identify at what session performance significantly improved above baseline measures. The statistical analysis and the use of confidence bands should provide converging evidence regarding learning effects in any cognitive task.

Because temporal stability confidence intervals, based on the standard error of measurement, can be constructed for all measures in all tests in the current battery, they will be used to evaluate the extent to which learning does or does not take place. This procedure will be used at the group level in collecting normative data, and at the individual level with the patients with stroke. The confidence intervals not only provide an index of learning, they also provide an indication as to the point where performance improves to a level where the changes can be attributed to the learning process.

#### **5.4 Experiment 5.1**

The primary aim of the current experiment is to collect normative data on the test battery, using a healthy, community-living sample that is more representative of the stroke population. This normative sample will be used to evaluate both the presence or absence of cognitive deficits among patients who have suffered a stroke, and the presence or absence of learning effects across the test sessions. A secondary aim is to confirm the psychometric properties of the test battery with this older sample. In short, the aim is to develop a psychometric “history” for all tasks using a sample across the age range during which 85% of strokes occur.

## 5.5 Method

### 5.5.1. Ethics

Approval for this body of research was sought and approved by the Australian Catholic University's Ethics Committee. All research reported in this chapter was collected in accordance with the conditions established under the approval number Q 2011 73.

### 5.5.2 Participants

Thirty participants were initially recruited for the study. Three participants were not able to complete the six sessions in the required two-week period. The data from these participants were not utilised. The remaining participants in the study were 27 volunteers (10 male and 17 female) whose ages ranged from 55 to 87 ( $M = 65.77$ ,  $SD = 8.13$ ). All participants indicated that they were in a current state of good health, and all lived independently in the community. One participant had suffered two Transient Ischaemic Attacks in the past, the most recent being more than 10 years ago. Both were resolved within 24 hours. None of the participants had a diagnosis of any condition that was associated with impaired cognition.

### 5.5.3 Materials and Procedure

The full battery as described in Chapter 3 and in Table 5.1 was administered across six sessions during a 2-week period. All tests were administered according to the descriptions in Chapter 3. In the Anagram and Snap tasks, the first two sessions were presented in single-task format. In Sessions 3 and 4 they were presented in the single-event detection dual task format, and in Sessions 5 and 6 they were presented in the two-event detection dual task format.

Table 5.1

*Tasks Administered in the Test Battery across Testing Sessions*

<b>Task</b>	<b>Condition</b>	<b>Sessions</b>					
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b><i>Snap</i></b>	Single Task	✓	✓				
	Dual Task						
	Single Event			✓	✓		
	Two Events					✓	✓
<b><i>Anagram</i></b>	Single Task	✓	✓				
	Dual Task						
	Single Event			✓	✓		
	Two Events					✓	✓
<b><i>Digit Span</i></b>							
<b><i>Forward</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Backward</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Stroop</i></b>							
<b><i>Read</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Colour</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Colour-word</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Verbal Fluency</i></b>							
<b><i>Phonemic</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Semantic</i></b>	Single	✓	✓	✓	✓	✓	✓
<b><i>Rey Tangled Lines. (RTL)</i></b>	Single	✓	✓	✓	✓	✓	✓

## 5.6 Results

### 5.6.1 Snap Task

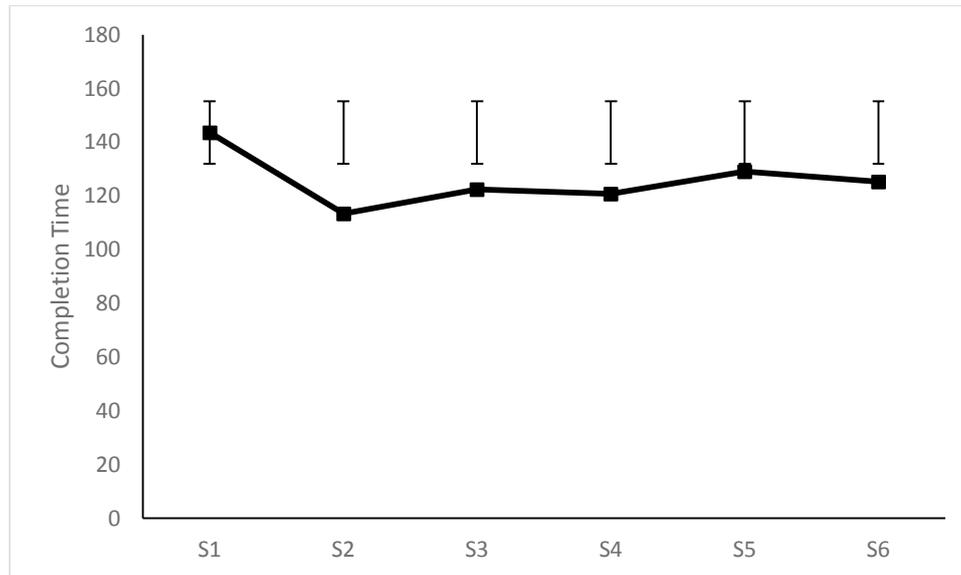
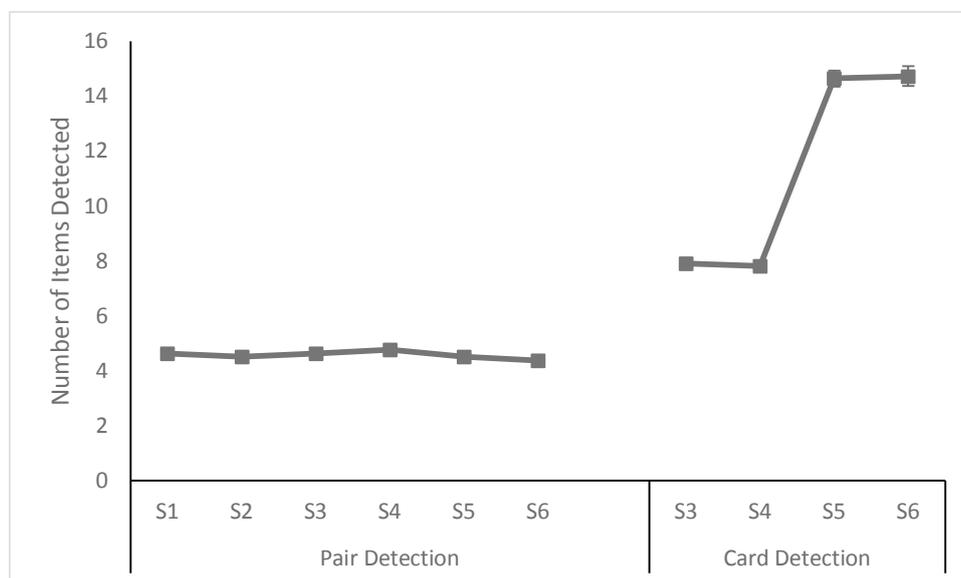


Figure 5.4. Performance on the Snap task as a function of testing session. Error bars represent temporal stability 95% confidence intervals based on initial test performance.

**5.6.1.1 Repetition.** The completion times on the Snap task are presented in Figure 5.4. A one-way ANOVA indicated that there were significant differences between sessions,  $F(5,130) = 5.06, p < .001, \eta_p^2 = .34$ . The improvement in performance did not contain a significant linear component,  $F(1,26) = .74, p = .40, \eta_p^2 = .07$ , but there were significant quadratic,  $F(1,26) = 7.15, p = .01, \eta_p^2 = .41$ , cubic,  $F(1,26) = 16.60, p < .001, \eta_p^2 = .62$ , and fifth order components,  $F(1,26) = 11.23, p = .003, \eta_p^2 = .53$ . There were significant differences between Session 1 and Sessions 2, 3, and 4, but not between Sessions 5 and 6 when the dual task became more difficult. The complex learning component analysis and the patterns of contrasts reflect changes in the dual task requirements across sessions. Compared to the single task data, there is a slowing with the introduction of a second task at Session 3. There was also a slowing with the transition from single-detection task at Session 4 to the dual-detection task at Session 5.

**5.6.1.2 Pair detection and dual task detection.** Accuracy in detecting the pair on the primary Snap task, and the accuracy in detecting the cards on the secondary task in the dual task situation are presented in Figure 5.5.



*Figure 5.5.* Pair detection (primary task) and card detection (secondary task) performance on the Snap task as a function of testing session. Error bars represent the standard error of the mean.

Pair detection was on ceiling with very few errors being made. Likewise on the card detection secondary task, performance was near perfect when a single card had to be detected, and performance was very close to ceiling when two cards had to be detected. In short, as far as detecting a pair on the primary task, or specific cards on the dual task detection component, participants were able to complete all aspects with high levels of proficiency.

**5.6.1.3 Slopes and intercepts.** Because there was no significant linear component in the main effect of testing sessions, slopes were not considered to be a good measure of learning. As a consequence, this measure was not applied in any subsequent evaluation of learning in the Snap task.

**5.6.1.4 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.4 are largely consistent with the outcomes of the ANOVA.

There was a significant improvement in performance after the first session, and performance remained outside the 95% confidence bands on all subsequent sessions, although performance was not far outside the lower limit in Sessions 5 and 6.

**5.6.1.5 Reliability.** The simple correlations between the testing sessions are presented in Table 5.2 and reliability estimates based on intraclass correlations are presented in Table 5.3. All bivariate correlations in this analysis and all subsequent ones were significant at the .05 level unless otherwise indicated. Intraclass correlations were based on a two-factor mixed model where consistency rather than absolute levels was examined. Again, all intraclass correlations were significant unless otherwise indicated.

Table 5.2

*Pearson's Correlation Coefficients for Six Testing Sessions on the Snap Task*

Session	Test Session				
	S1	S2	S3	S4	S5
S2	0.65				
S3	0.61	0.82			
S4	0.60	0.81	0.97		
S5	0.61	0.82	0.84	0.81	
S6	0.58	0.80	0.86	0.81	0.99

Note: all correlations significant at .05 level

Table 5.3

*Intraclass Correlations [95% Confidence Intervals] for the Snap Task*

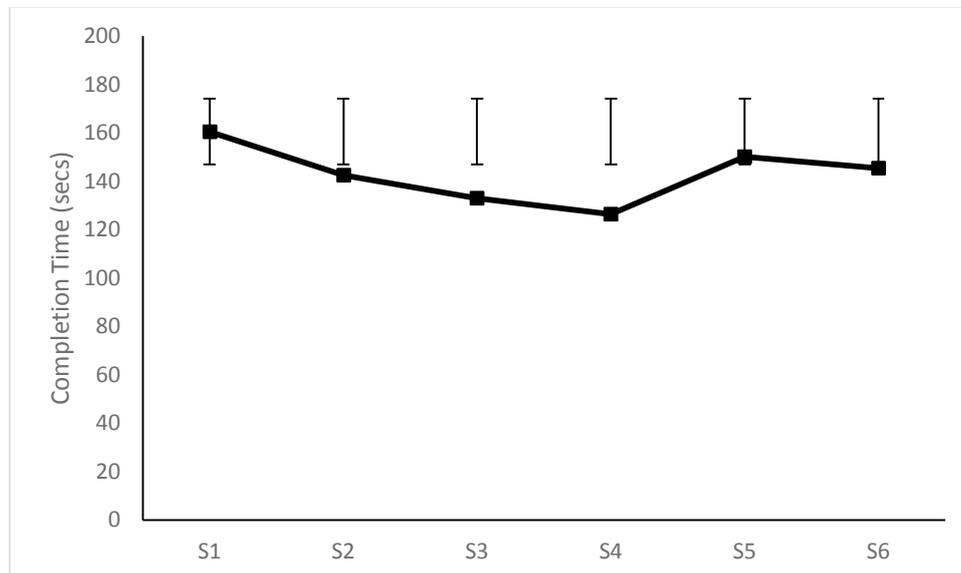
Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Snap	.61 [.29, .80]	.81 [.63, .91]	.96 [.92, .98]	.80 [.59, .90]	.98 [.95, .99]

Note: All correlations significant at .05 level

It is apparent in all measures that performance was more variable on the first test than on the others. That is, by the second test performance appeared to have stabilised. Performance across the remaining five sessions was highly reliable. It was the case, however, that reliability was better between sessions where there was no change in dual task requirement (Sessions 3 and 4, Sessions 5 and 6) than when there was a change in requirements (Sessions 2 and 3, Sessions 4 and 5).

### 5.6.2 Anagram Task

**5.6.2.1 Repetition.** The scores for completion time are presented in Figure 5.6 The Figure suggests that there was an improvement in completion time across the first four sessions, however with the introduction of two-event detection there was a slowing in the Anagram task, indicating that the dual task manipulation was successful.



*Figure 5.6.* Completion time on the Anagram task as a function of testing session. Error bars represent 95% confidence interval for repeated testing.

A one-way ANOVA indicated that there were significant differences between sessions,  $F(5,130) = 10.79$ ,  $p < .001$ ,  $\eta_p^2 = .37$ . There were significant differences between

Session 1 and all subsequent sessions except for Session 5 when the dual task became more difficult. The improvement in performance did not contain a significant linear component,  $F(1,26) = 2.88, p = .10, \eta_p^2 = .10$ , but there were significant quadratic,  $F(1,26) = 22.40, p < .001, \eta_p^2 = .47$ , cubic,  $F(1,26) = 7.33, p = .012, \eta_p^2 = .23$ , fourth order,  $F(1,26) = 6.73, p = .016, \eta_p^2 = .21$  and fifth order components,  $F(1,26) = 8.03, p = .009, \eta_p^2 = .24$ .

**5.6.2.2 Anagram errors and dual task detection.** The number of errors on the Anagram task is presented in the left hand panel of Figure 5.7. While there was a significant difference among the testing sessions,  $F(1,125) = 4.59, p = .001, \eta_p^2 = .16$ , this was attributable to the spike in errors on Session 2. There were no significant differences between Session 1 and any other testing session.

Performance of the category detection task was on ceiling when only a single event had to be detected in Sessions 3 and 4, but was off ceiling in Sessions 5 and 6 when two categories had to be monitored. However, there was a significant improvement on the task on Session 6,  $t(25) = 2.11, p = .04$ .

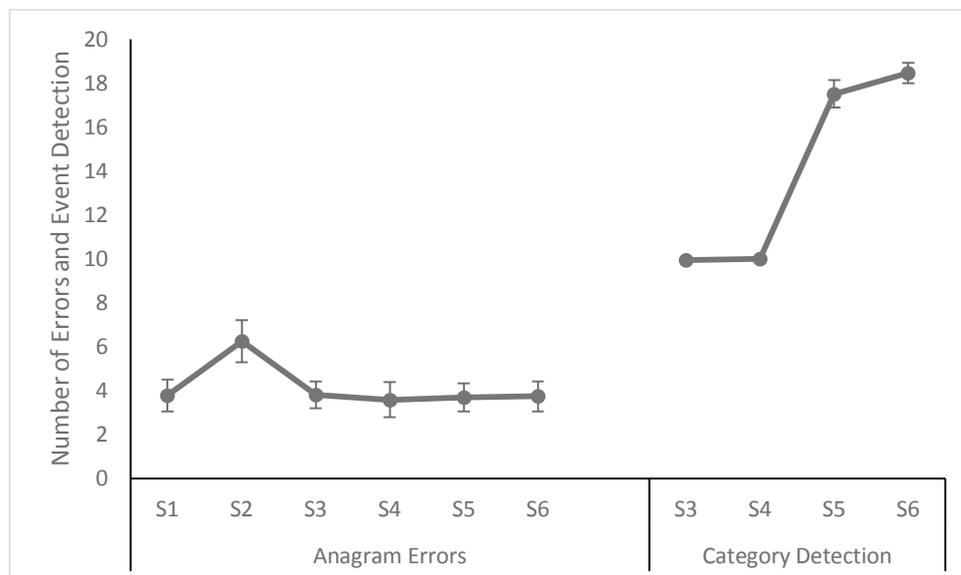


Figure 5.7. Number of errors and number of dual task detections on the Anagram task as a function of testing session. Error bars represent standard error of the mean.

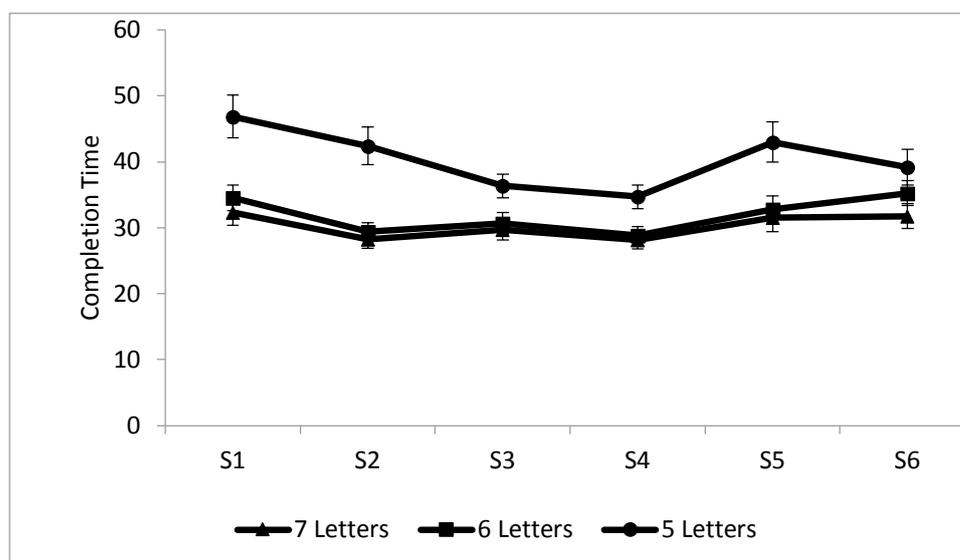


Figure 5.8. Completion time on the Anagram task as a function of testing session and letter length. Error bars represent the standard error of the mean.

**5.6.2.3 Repetition and Benchmarks.** The effects of anagram length on completion time are summarised in Figure 5.8. These data were subjected to a 6 session by 3 letter length repeated measures ANOVA. There were significant effects of letter length,  $F(2,52) = 32.24$ ,  $p < .001$ ,  $\eta_p^2 = .56$ . There was a significant difference between 7-letter completion times and 6-letter completion times,  $F(1,26) = 7.27$ ,  $p = .012$ ,  $\eta_p^2 = .23$ , and 6-letter anagrams were completed significantly faster than 5-letter anagrams,  $F(1,26) = 33.57$ ,  $p < .001$ ,  $\eta_p^2 = .57$ .

There were also significant differences across testing sessions,  $F(5,130) = 10.11$ ,  $p < .001$ ,  $\eta_p^2 = .27$ . There was no significant linear component,  $F(1,26) = 0.75$ ,  $p = .39$ ,  $\eta_p^2 = .03$ , but there were significant quadratic,  $F(1,26) = 25.09$ ,  $p < .001$ ,  $\eta_p^2 = .50$ , cubic,  $F(1,26) = 5.78$ ,  $p = .024$ ,  $\eta_p^2 = .19$ , and fifth order components,  $F(1,26) = 9.33$ ,  $p = .005$ ,  $\eta_p^2 = .27$ .

The interaction between length and session was also significant,  $F(10,250) = 4.53$ ,  $p < .001$ ,  $\eta_p^2 = .15$ . As is evident in Figure 5.8 length differences were more apparent in the early testing sessions than in the latter ones, due to greater improvements in solution times for 5-letter anagrams than the longer anagrams. However, the 5-letter anagrams were more vulnerable to the two-detection version of the dual task than were the 6-letter and 7-letter

anagrams. The finding that 5-letter anagrams were significantly slower to complete than 6-letter and 7-letter confirms the manipulation of task difficulty and is thus consistent with benchmark expectations.

**5.6.2.4 Slopes and intercepts.** Again because there was no significant linear component in the main effect for testing session, the use of slopes as a measure of learning was deemed inappropriate. As was the case with the snap task, this measure is not used in subsequent evaluations of learning in this task.

**5.6.2.5 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.6 confirmed the outcomes of the ANOVA. There was a significant improvement in performance after the first session, and performance remained outside the 95% confidence bands on all subsequent sessions, except for Session 5 when the two-event dual task condition was first implemented.

**5.6.2.6 Reliability.** Simple and intraclass correlations between the testing sessions are presented in Tables 5.4 and 5.5 respectively. Internal consistency measures for each session, based on completion times for 7-letter words, 6-letter words and 5-letter words, are presented in the first row of Table 5.4. The reliability estimates on the first test session are marginal, but by the third session all measures indicate that the test has acceptable reliability.

Table 5.4

*Pearson's Correlation Coefficients for Six Testing Sessions on the Anagram Task*

Sessions	S1	S2	S3	S4	S5	S6
Alpha	.65	.68	.88	.85	.88	.85
S2	0.63					
S3	0.79	0.84				
S4	0.72	0.85	0.87			
S5	0.66	0.80	0.87	0.88		
S6	0.83	0.72	0.86	0.82	0.87	

Note: All correlations significant at .05 level.

Table 5.5

*Intraclass Correlations [95% Confidence Intervals] for the Anagram Task*

Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Anagram	.62 [.31, .81]	.83 [.65, .92]	.87 [.72, .98]	.77 [.55-.98]	.86 [.78, .93]

Note: All correlations significant at .05 level.

### 5.6.3 Stroop Colour Word Task

**5.6.3.1 Repetition.** The number of items completed in 45 seconds on the three different components of the Stroop task are summarised in Figure 5.9. The Figures clearly indicate strong learning effects in all measures.

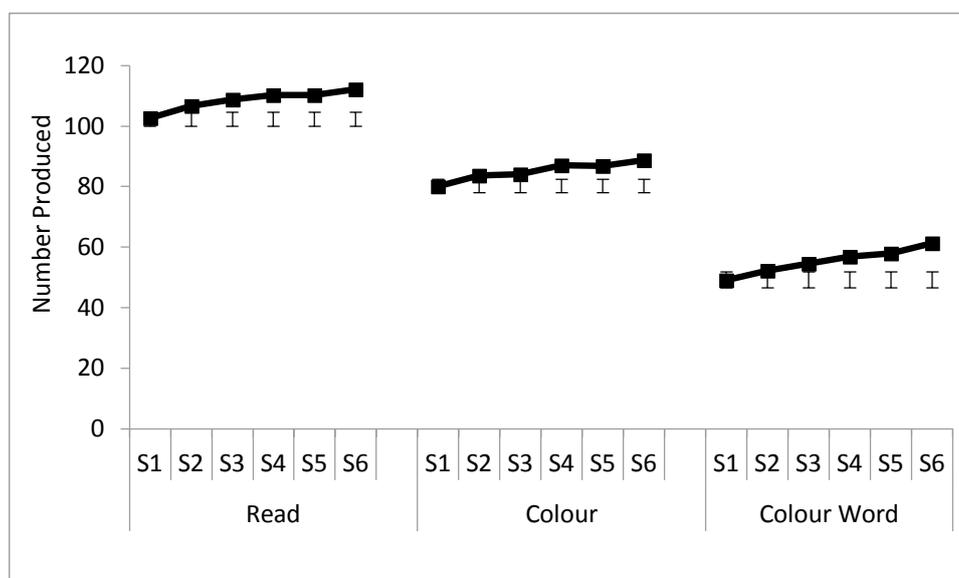


Figure 5.9. Performance on the Stroop task as a function of test sessions.

The read, colour naming, and colour-word naming scores were all analysed via one-way repeated measures ANOVA. For the word reading measure there were significant differences across testing sessions,  $F(5,130) = 9.37, p < .001, \eta_p^2 = .22$ . There was a strong linear component to the changes in performance across sessions,  $F(5,130) = 16.30, p < .001, \eta_p^2 = .39$ . Post-hoc analyses indicated that by the second session there was a significant improvement compared to the first session. The differences between Sessions 1 and the remaining sessions were also statistically significant. In addition to the linear component there was a significant cubic component,  $F(5,130) = 4.45, p = .045, \eta_p^2 = .15$ .

For the colour naming scores there was again a significant effect of sessions,  $F(5,130) = 11.15, p < .001, \eta_p^2 = .31$ , with a strong linear component as the only significant contributor to change,  $F(5,130) = 17.87, p < .001, \eta_p^2 = .52$ . Again, post-hoc analyses indicated a significant difference between baseline performance and all subsequent sessions.

The same pattern occurred with colour-naming scores with a significant effect of test sessions,  $F(5,130) = 26.74, p < .001, \eta_p^2 = .52$ , characterised by a significant linear component,  $F(5,130) = 78.39, p < .001, \eta_p^2 = .76$ . Again, there was a significant difference

above baseline by the second session and this improvement was maintained across subsequent sessions.

**5.6.3.2 Slopes and intercepts.** For all three measures the best-fitting linear regression line was estimated, and subsequent slope values and intercepts were derived for each participant. For the read measure, the correlation between slopes and intercepts was statistically significant,  $r = -.41, p = .039$ . For the colour naming and colour-word naming, the correlations were not significant,  $r = .10, p = .62$ , and,  $r = .05, p = .78$ .

**5.6.3.3 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.9 indicate that there was a significant improvement in performance after the first session on the three measures, and performance remained outside the 95% confidence bands on all subsequent sessions.

**5.6.3.4 Reliability.** Reliability information by way of intraclass correlations for the three measures of the Stroop task is presented in Tables 5.6. Reliability estimates were excellent for all test sessions for all three sub-tests.

Table 5.6

*Intraclass Correlations [95% Confidence Intervals] for the sub-tests of the Stroop task*

Measure	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Read	.87 [.73, .93]	.80 [.59, .90]	.90 [.78, .95]	.89 [.76, .94]	.95 [.88, .97]
Colour	.84 [.68, .92]	.93 [.84, .96]	.91 [.81, .96]	.94 [.87, .97]	.93 [.84, .96]
Colour-Word	.93 [.85, .96]	.94 [.88, .97]	.90 [.78, .95]	.90 [.78, .95]	.97 [.94, .98]

Note: All correlations significant at .05 level.

The correlations among testing sessions for the three measures are presented in Tables 5.7 to 5.9. Again, there were strong correlations across all sessions across all sub-tests.

Table 5.7

*Pearson's Correlation Coefficients for Six Testing Sessions on the Read Sub-test of the Stroop task*

Session	S1	S2	S3	S4	S5
S2	0.87				
S3	0.62	0.80			
S4	0.60	0.80	0.90		
S5	0.65	0.80	0.86	0.89	
S6	0.61	0.77	0.81	0.92	0.95

Note: All correlations significant at .05 level

Table 5.8

*Pearson's Correlation Coefficients for Six Testing Sessions on the Colour Naming Sub-test of the Stroop task*

Session	S1	S2	S3	S4	S5
S2	0.87				
S3	0.86	0.93			
S4	0.85	0.92	0.92		
S5	0.86	0.90	0.93	0.94	
S6	0.83	0.92	0.88	0.92	0.95

Note: All correlations significant at .05 level.

Table 5.9

*Pearson's Correlation Coefficients for Six Testing Sessions on the Colour-Word Sub-test of the Stroop task*

Session	S1	S2	S3	S4	S5
S2	0.93				
S3	0.90	0.95			
S4	0.84	0.91	0.92		
S5	0.88	0.92	0.92	0.90	
S6	0.90	0.95	0.93	0.92	0.98

Note: All correlations significant at .05 level.

#### 5.6.4 Digit Span

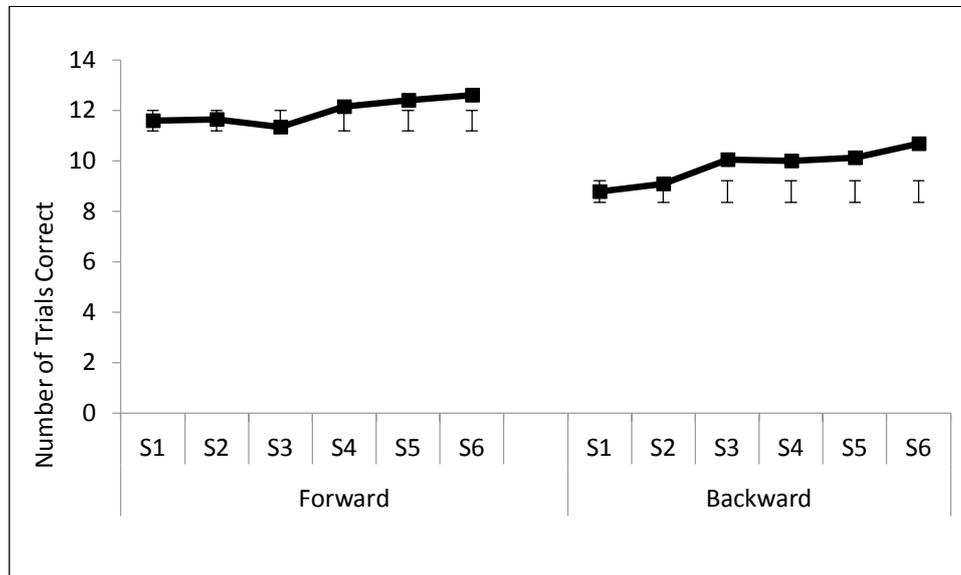


Figure 5.10. Performance on Digit Span task across test sessions.

**5.6.4.1 Repetition and benchmarks.** Performance on the Digit Span task is summarised in Figure 5.10. A repeated measures ANOVA with 2 directions and 6 testing sessions as the factors, revealed that digit recall was significantly better in the forward direction than the backward direction,  $F(1,26) = 60.84, p < .001, \eta_p^2 = .70$ . There was a

significant effect of session,  $F(5,130) = 12.07, p < .001, \eta_p^2 = .32$ , of which only the linear component was significant,  $F(1,26) = 38.87, p < .001, \eta_p^2 = .60$ . The interaction between direction and test sessions was not significant,  $F(5,130) = 2.17, p = .06, \eta_p^2 = .08$ .

In forward recall, the linear component was the only significant contributor to changes across sessions,  $F(1,26) = 14.10, p = .001, \eta_p^2 = .15$ , but significant improvement over Session 1 did not emerge until Session 5. In backward recall, the linear component was the only significant contributor,  $F(1,26) = 23.32, p < .001, \eta_p^2 = .47$ , and the improvement was significant by Session 3.

**5.6.4.2. Slopes and intercepts.** For both forward span and backward span the correlations between slopes and intercepts were statistically significant,  $r = -.58, p < .001$  and  $r = -.59, p < .001$ , suggesting that those who showed the most improvement were those who were initially the poorest on the task. Therefore, slopes in this task may not provide a good estimate of the degree of learning involved in the task.

**5.6.4.3 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.10 confirmed the outcomes of the ANOVA. With forward span, the effects of repetition only emerged at Session 5, but the advantage was maintained at Session 6. Learning was more robust in backward span in that performance was first observed outside the 95% confidence bands on Session 3, and remained outside the bands on the subsequent sessions.

**5.6.4.4 Reliability.** Reliability estimates for forward and backward digit span are presented in Tables 5.10. The intraclass correlations were stronger for forward recall than backward recall, but in both cases the reliability estimates were low in comparison to the other tests used in the battery.

Table 5.10

*Intraclass Correlations [95% Confidence Intervals] for Forward and Backward Digit Span*

Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
Forward	.71 [.45, .85]	.70 [.43, .84]	.57 [.24, .77]	.70 [.43, .84]	.71 [.45, .85]
Backward	.68 [.41, .84]	.68 [.41, .84]	.66 [.38, .82]	.65 [.37, .82]	.87 [.72, .93]

Note: All correlations significant at .05 level.

The correlations between testing sessions are presented in Tables 5.11 and 5.12.

Again, it is apparent that performance on the span tasks across sessions was more variable than was the case on other tests.

Table 5.11

*Pearson's Correlation Coefficients for Six Testing Sessions on Forward Span*

Session	S1	S2	S3	S4	S5
S2	0.71				
S3	0.67	0.70			
S4	0.58	0.72	0.57		
S5	0.69	0.77	0.75	0.71	
S6	0.56	0.65	0.74	0.66	0.71

Note: All correlations significant at .05 level.

Table 5.12

*Pearson’s Correlation Coefficients for Six Testing Sessions on Backward Span*

Sessions	S1	S2	S3	S4	S5
S2	0.69				
S3	0.71	0.69			
S4	0.67	0.44	0.66		
S5	0.70	0.57	0.78	0.66	
S6	0.65	0.44	0.79	0.75	0.87

Note: All correlations significant at .05 level.

**5.6.5 Rey Tangled Lines**

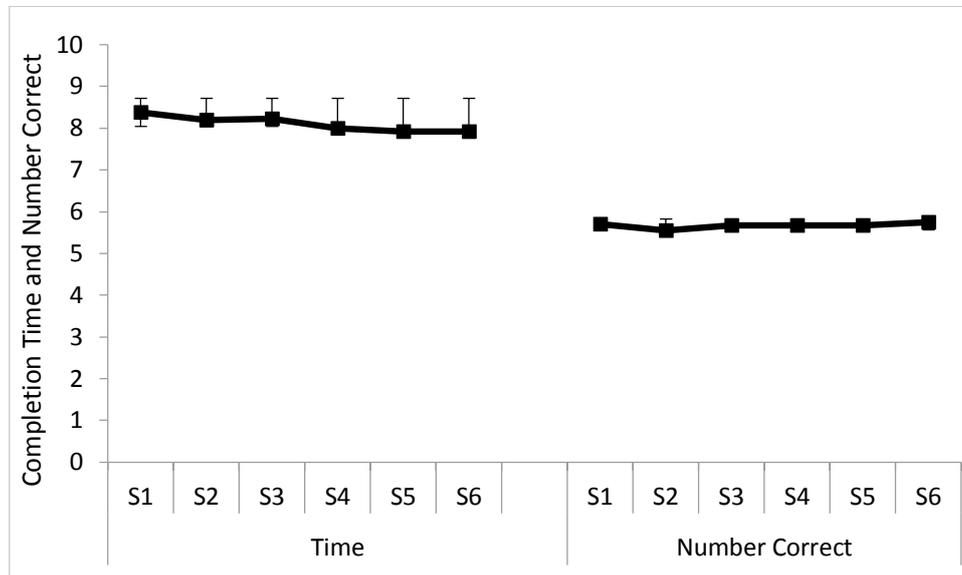


Figure 5.11. Performance on the Rey Tangled Lines test as a function of test session.

**5.6.5.1 Repetition.** Figure 5.11 summarises the data from the Rey Tangled Lines test

A one-way repeated measures ANOVA indicated that there was a significant difference between test sessions on completion time,  $F(5,130) = 2.62, p = .026, \eta_p^2 = .09$ . The linear component of the difference across scores was the only significant effect,  $F(1,26) = 8.79, p =$

.006,  $\eta_p^2 = .25$ . Post-hoc analyses indicated that performance significantly differed from baseline levels only at Sessions 5 and 6.

The number of correct responses is presented in the right hand panel of Figure 5.11. Performance was basically on ceiling, with very few errors being made. Consequently, there were no significant differences between testing sessions,  $F(5,130) = .35, p = .88, \eta_p^2 = .01$ .

**5.6.5.2 Slopes and intercepts.** The correlation between slopes and intercepts was not statistically significant,  $r = -.26, p = .18$ .

**5.6.5.3 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.11 indicated that significant improvement in performance was detected only on Sessions 5 and 6.

**5.6.5.4 Reliability.** Reliability data are presented in Tables 5.13 and 5.14. The intraclass correlations indicated good levels of reliability. Internal consistency was evaluated using Chronbach's alpha and was based on the six trials that make up the total score. The values are presented in the first row of Table 5.15, and represent adequate levels of internal consistency.

Table 5.13

*Intraclass Correlations [95% Confidence Intervals] for Rey Tangled Lines Test*

Task	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
RTL	.79 [.58, .89]	.81 [.63, .91]	.75 [.52, .87]	.83 [.66, .92]	.90 [.79, .95]

Note: All correlations significant at .05 level.

Table 5.14

*Internal Consistency and Pearson's Correlation Coefficients for Six Testing Sessions on the Rey Tangled Lines test*

Sessions	S1	S2	S3	S4	S5	S6
Alpha	.83	.77	.75	.88	.83	.85
S2	0.79					
S3	0.74	0.81				
S4	0.73	0.85	0.76			
S5	0.86	0.78	0.78	0.85		
S6	0.77	0.83	0.75	0.93	0.90	

Note: All correlations significant at .05 level.

Chronbach's Alpha, simple correlations and intraclass correlations complement one another in showing that the Rey Tangled Lines Test has excellent reliability.

### **5.6.6 Phonemic Fluency Test**

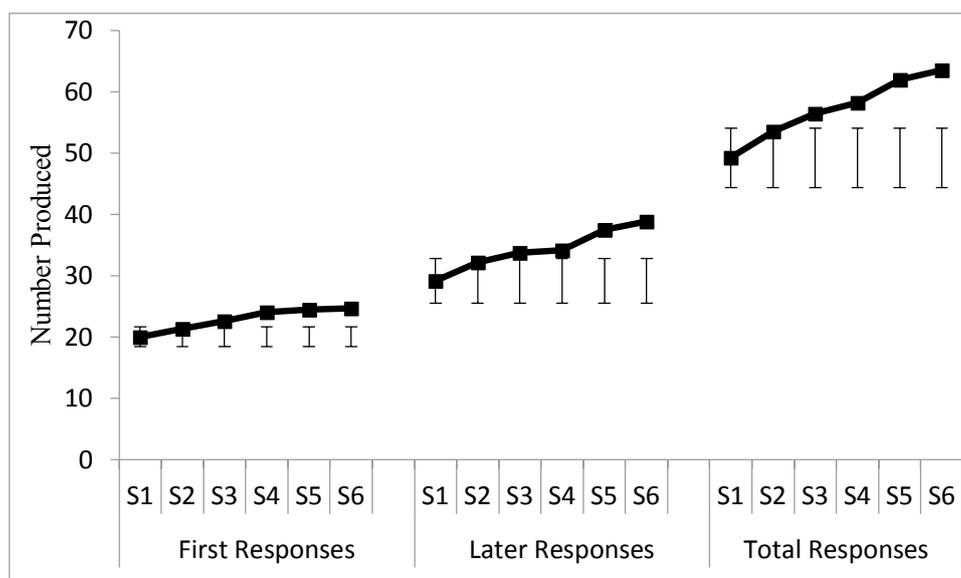


Figure 5.12. Performance on the phonemic fluency task as a function of test sessions.

**5.6.6.1 Repetition.** The three measures derived from the phonemic fluency test are summarised in Figure 5.12. As the total score is a composite of the other two measures it was not sensible to conduct factorial ANOVAs on the data. Thus, three separate one-way repeated measures analyses were conducted on each of the three dependent measures.

For the first responses data, there was a significant difference between test sessions,  $F(5,130) = 15.26, p < .001, \eta_p^2 = .37$ . The linear component made a significant contribution to these differences,  $F(1,26) = 53.55, p < .001, \eta_p^2 = .67$ , as did the quadratic component,  $F(1,26) = 5.10, p = .032, \eta_p^2 = .67$ . Significant repetition effects emerged by the third session.

For the later responses, there was also a significant difference between test sessions,  $F(5,130) = 10.97, p < .001, \eta_p^2 = .29$ . Only the linear component made a significant contribution,  $F(1,26) = 38.34, p < .001, \eta_p^2 = .60$ . Improvements in performance were significantly above baseline by the second session.

For the total scores, there was again a significant difference between test sessions,  $F(5,130) = 20.91, p < .001, \eta_p^2 = .45$ , with the linear component only making a significant contribution,  $F(1,26) = 73.10, p < .001, \eta_p^2 = .74$ . Again, learning was significantly higher than for baseline by the second session.

Prior research has shown that most people show a “fast start slow finish” pattern of performance leading to the expectation that the number of items produced in the first 15-second response period should produce more items than any single subsequent 15-second period (Fernaesus & Almkvist, 1998). To check this benchmark, performance on the first 15-second period was compared to the average number produced in each of the remaining three 15-second periods. On average, participants produced a combined total of 22.94 items in the first period compared to 11.17 items in each of the remaining periods,  $F(1,26) = 324.16, p < .001, \eta_p^2 = .93$ .

**5.6.6.2 Slopes and intercepts.** The correlations between slopes and intercepts were not statistically significant for first responses,  $r = -.26, p = .18$ , later responses,  $r = -.17, p = .37$ , or total scores,  $r = -.10, p = .61$ .

**5.6.6.3 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.12 were consistent with the outcomes of the ANOVA. There was a significant improvement in performance after the third session on the first responses, and this improvement was maintained on all subsequent sessions. In the later responses and the total scores, significant improvements had emerged by Session 2 and were maintained thereafter.

**5.6.6.4 Reliability.** The intraclass correlations on the three measures are presented in Table 5.15. The reliability of all three measures was excellent.

Table 5.15

*Intraclass Correlations [95% Confidence Intervals] for the Three Scores on the Phonemic Fluency Task*

Measures	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
First Responses	.71 [.46, .86]	.84 [.68, .92]	.80 [.61, .90]	.80 [.61, .90]	.85 [.69, .92]
Later Responses	.84 [.68, .92]	.73 [.48, .86]	.79 [.59, .90]	.85 [.69, .92]	.85 [.69, .92]
Totals	.91 [.82, .96]	.82 [.64, .91]	.86 [.71, .93]	.89 [.77, .95]	.89 [.77, .95]

Note: All correlations significant at .05 level.

The correlations among testing sessions for the three measures are presented in Tables 5.16 to 5.18. Again, there were strong correlations across all sessions for all three measures.

Table 5.16

*Pearson's Correlation Coefficients for First Responses on Phonemic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.72				
S3	0.72	0.84			
S4	0.71	0.71	0.81		
S5	0.74	0.84	0.90	0.81	
S6	0.69	0.80	0.84	0.82	0.85

Note: All correlations significant at .05 level.

Table 5.17

*Pearson's Correlation Coefficients for Six Testing Sessions on the Later Responses in the Phonemic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.84				
S3	0.80	0.74			
S4	0.77	0.72	0.80		

S5	0.80	0.72	0.84	0.86	
S6	0.86	0.73	0.87	0.80	0.85

---

Note: All correlations significant at .05 level.

Table 5.18

*Pearson's Correlation Coefficients for Six Testing Sessions on the Total Responses in the Phonemic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.92				
S3	0.85	0.83			
S4	0.82	0.85	0.86		
S5	0.83	0.84	0.90	0.90	
S6	0.89	0.83	0.90	0.84	0.89

Note: All correlations significant at .05 level.

**5.6.7 Semantic Fluency Test**

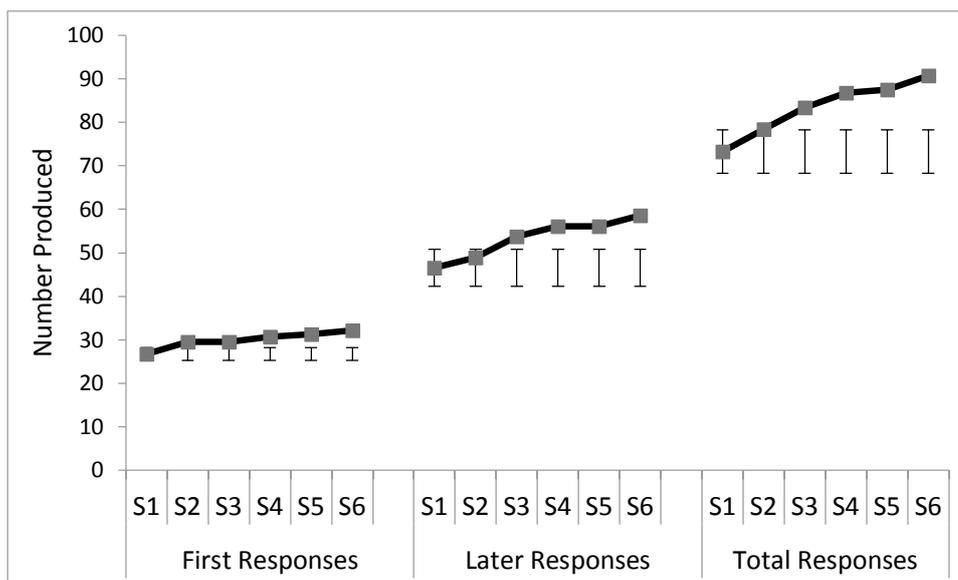


Figure 5.13. Performance on the semantic fluency task as a function of test sessions.

**5.6.7.1 Repetition.** Figure 5.13 presents the data from the three measures on the semantic fluency task. As was the case with phonemic fluency, and for the same reasons, three separate one-way repeated measures were conducted on each of the three dependent measures. For all three measures there were significant differences between test sessions, *F*

(5,130) = 5.94,  $p < .001$ ,  $\eta_p^2 = .19$ ,  $F(5,130) = 13.59$ ,  $p < .001$ ,  $\eta_p^2 = .34$ , and,  $F(5,130) = 24.82$ ,  $p < .001$ ,  $\eta_p^2 = .49$ , for first responses, later responses and total scores respectively.

For all three measures the linear component was the only component to make a significant contribution to the changes,  $F(1,26) = 33.66$ ,  $p < .001$ ,  $\eta_p^2 = .56$ ,  $F(1,26) = 50.25$ ,  $p < .001$ ,  $\eta_p^2 = .65$ , and,  $F(1,26) = 102.34$ ,  $p < .001$ ,  $\eta_p^2 = .79$ , for first, later and total scores respectively. For all three scores it took three sessions for performance to improve significantly above baseline.

Analysis of the “fast start slow finish” characteristics of the task confirmed that, on average, participants produced a combined total of 29.47 items in the first period compared to 18.14 items in each of the remaining periods,  $F(1,26) = 237.90$ ,  $p < .001$ ,  $\eta_p^2 = .90$ .

**5.6.7.2 Slopes and intercepts.** The correlations between slopes and intercepts were not statistically significant for first responses,  $r = -.11$ ,  $p = .58$ , later responses,  $r = -.10$ ,  $p = .59$ , or total scores,  $r = -.01$ ,  $p = .94$ .

**5.6.7.3 Temporal stability confidence intervals.** The temporal stability confidence intervals presented in Figure 5.13 were consistent with the outcomes of the ANOVA. On all three measures, significant improvements in performance emerged by the third session. Performance remained outside the 95% confidence bands on all subsequent sessions for all three measures.

**5.6.7.4 Reliability.** The intraclass correlations on the three measures are presented in Table 5.19. The reliability of the first two repetitions on the first responses did not reach the desired level of reliability. However, reliability improved in the later testing sessions for this measure. For the later responses and the total scores, reliability was acceptable at all sessions.

Table 5.19

*Intraclass Correlations for the Three Scores on the Semantic Fluency Task*

Measures	S1 & S2	S2 & S3	S3 & S4	S4 & S5	S5 & S6
First Responses	.55 [.21, .76]	.47 [.11, .71]	.79 [.58, .89]	.85 [.69, .92]	.77 [.56, .89]
Later Responses	.73 [.49, .86]	.76 [.53, .88]	.83 [.66, .91]	.85 [.69, .92]	.89 [.77, .94]
Totals	.80 [.60, .90]	.82 [.64, .91]	.92 [.82, .96]	.94 [.87, .97]	.94 [.87, .97]

Note: All correlations significant at the .05 level.

The correlations among testing sessions for the three measures are presented in Tables 5.20 to 5.22. Again, there were strong correlations across all sessions for the three measures.

Table 5.20

*Pearson's Correlation Coefficients for First Responses on Semantic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.60				
S3	0.59	0.47			
S4	0.65	0.53	0.79		
S5	0.68	0.72	0.81	0.85	
S6	0.66	0.73	0.59	0.67	0.78

Note: All correlations significant at the .05 level.

Table 5.21

*Pearson's Correlation Coefficients for Six Testing Sessions on the Later Responses in the Semantic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.75				
S3	0.81	0.78			
S4	0.71	0.79	0.84		
S5	0.84	0.68	0.82	0.85	
S6	0.81	0.84	0.82	0.88	0.89

Note: All correlations significant at the .05 level.

Table 5.22

*Pearson's Correlation Coefficients for Six Testing Sessions on the Total Responses in the Semantic Fluency Task*

Sessions	S1	S2	S3	S4	S5
S2	0.82				
S3	0.92	0.84			
S4	0.87	0.82	0.92		
S5	0.89	0.83	0.89	0.94	
S6	0.91	0.88	0.91	0.92	0.94

Note: All correlations significant at the .05 level.

## 5.7 Discussion

The aims of Experiment 5.1 were threefold. The first aim was to provide normative data on each of the tests used in the battery in terms of both absolute levels of performance and the degree to which performance improved with repeated testing. The second aim was to establish that known benchmarks were present in the current samples. The third aim was to examine the psychometric properties of the tests in order to assess the likelihood that

meaningful results would emerge when applied to the patients with stroke. The results were largely positive on all counts.

### **5.7.1 Repetition Effects**

Repetition effects were evaluated in four different ways using the primary measures of each task. The presence of significant main effect for testing sessions in each ANOVA served as initial evidence that learning was taking place. All measures on all tasks met this criterion. However, while significant main effects were indicative of changes in performance across sessions, two follow-up analyses based on the ANOVA results explored the characteristics of the change. The strongest indicator of improvement was the extent to which improvement occurred in a linear fashion. With the exception of the Snap and Anagram tasks, there were strong linear components to the learning curve in all other primary measures. For the remaining tests, in all instances except for the initial responses in the phonemic fluency task, a significant main effect for testing sessions was all that was needed to explain the changes in scores across sessions. For the initial responses on the phonemic fluency task, a quadratic component was also required to satisfactorily account for learning. However, the linear component explained substantially more variance than the higher order factor. Across all the measures where dual task versions were not included, learning can be adequately described in terms of constant increases in performance each time the test is given.

The two instances where learning could not be described in terms of a linear component, namely the Snap and the Anagram tasks, were the tasks in which secondary were progressively included in addition to the single-task component. The data would suggest that the introduction of a second task caused disruption to the primary task and as a result the test session learning curve was complex. The absence of a significant linear component to the regression equation suggests that for the Snap and Anagram tasks, the linear slope is not an adequate measure of learning.

Post-hoc comparisons of the ANOVA results served as the second measure of learning. The intent here was to determine at which point in the testing regime learning could be detected. Consequently, performance on the first test session was compared to all subsequent sessions. In the cases where there was a linear component to the learning curve, significant improvements above baseline were most frequently observed at Sessions 2 or Session 3. In the case of digits forward and the Rey Tangled Lines test, the improvement did not emerge until Session 5. Thus, for the majority of measures, significant improvements on the test were apparent very early in the testing process.

While the significant linear components were suggestive that learning was occurring in an orderly fashion, the relationships between learning slopes and intercepts were examined to determine that the learning slopes were not just an artefact of initial levels of performance. In the case of digit forwards, digit backwards and for the read measure of the Stroop task, there was a significant correlation between slopes and intercepts. In each case those who were good at the task in the initial sessions did not improve much across sessions whereas those who were initially poor at the task did show marked improvement across sessions. This is an outcome that one might expect as it suggests that learning was most effective for those least skilled. However, it is also the case that regression to the mean would produce the same effect. Consequently, care needs to be taken in the interpretation of any linear changes in these measures.

The fourth measure was based on the psychometric properties of each test. Based on the standard error of measurement for the initial measure in each test, a temporal stability 95% confidence interval was calculated for each measure on the assumption that if learning was not occurring across sessions then the means of successive sessions should lie within the confidence intervals. If the mean was outside those confidence intervals then the assumption was made that an additional source of variance was producing this change rather than error variance. Using this measure, the outcomes were identical to those obtained via planned

comparisons with the initial testing session. For most measures the mean fell outside the 95% confidence interval at Sessions 2 or 3, and at Session 5 for digits forward and Rey Tangled Lines. Of the four measures used to evaluate learning outcomes, this is probably the most useful in that these same confidence intervals can be applied to individual data to establish if and when a change in performance can be attributed to learning.

### **5.7.2 Dual-Task and Benchmark Effects.**

There were a number of benchmark effects that were explored in the different tasks. The first involved the introduction of dual tasks effects in the snap and Anagram tasks. In the Snap task completion time slowed when the single-event dual task was incorporated for the first time in Session 3 and slowed again when the two-event dual task was administered for the first time in Session 5. However, participants rapidly adapted to the dual task manipulations in that the following sessions, Sessions 4 and 6, completion time improved once again. Performance on the detection task in both instances was near ceiling.

In the case of the Anagram task, the single-event detection appeared to place minimal processing demands on participants. However, in the case of the two-event detection condition, there was an appreciable slowing in completion time and detection of the two events was also adversely affected. In short, in both tasks dual task effects can be observed in the data with the two-event condition being more disruptive than the single-event detection.

There were other benchmark effects in the various tasks. In the Anagram task it was expected that solution time should differ as a degree of the number of letters in the anagram that remained intact, with solution times being faster for 7-letter anagrams, than 6-letter anagrams and 5-letter anagrams. There were significant differences in completion time for the different anagram lengths, with 5-letter anagrams being much harder to solve than either the 7-letter and 6-letter anagrams. Importantly, the effect of anagram length differences was attenuated as participants had more practice on the task.

In the digit span it was expected that span scores for forward span would be higher than those for backward span. This expectation was met.

It was also expected that there would be differences among the three components of the Stroop task. Again the data conformed to expectations in that more items were produced on the read sub-task than the colour naming sub-task and fewest were recalled on the colour-word component.

Lastly, prior research had indicated that on both the phonological and semantic fluency tasks more items were generated in the early phases of the task than later phases (Fernaes & Almkvist, 1998). The current data showed just such a pattern. In both tasks the majority of words produced were produced in the first 15-second segment. Subsequent 15-second segments contained numerous pauses while attempts were made to generate new items. In fact, while there was substantial improvement across sessions in both forms of the fluency task, the proportion of the total responses that were generated in the first 15-second response period remained constant across testing sessions, with .40 in the phonemic fluency task and .35 for the semantic fluency task. The difference in proportions is attributable to many more items being generated in the latter responses in the semantic fluency tasks than in the phonemic fluency task, as is evident in Figures 5.11 and 5.12.

### **5.7.3 Psychometric Properties**

In looking at the psychometric properties of the tests, a number of generalised comments can be made. Firstly, measures of reliability were generally weakest on the first testing occasion and were stronger on subsequent sessions. Secondly, the values of the obtained measures were in almost all instances in the acceptable to very strong range. The exceptions to this rule involved the two digit span measures and the first responses in the semantic fluency task.

The current experiment replicates the high reliability estimates for the Snap and Anagram tasks that were found in Experiment 4.2 and shows that the Digit Span task, like the

Immediate Serial Recall task in Experiment 4.2, has the least consistent reliability of any of the tasks in the battery. However, for the other tasks in the battery there are no reliability issues that would compromise understanding of changes in performance with repeated testing.

#### **5.7.4 Demographic Effects**

Absolute levels of performance on many neuropsychological and cognitive tests can be influenced by demographic factors such as age, education, gender, and socio-economic status. While gender effects typically have little influence on performance (Strauss et al., 2006), age and education can result in substantial changes in performance, and consequently the norms for many neuropsychological tests are broken down into age and education bands.

The demographic characteristics of the normative sample were explored by a series of 2 x 6 ANOVAs in which gender, age and education effects were explored. The presence of a main effect for any of the demographic variables would indicate that the normative sample as a whole might not be the appropriate comparison group with which to evaluate any stroke patient. However, the crucial outcome of these analyses involved the interaction between demographic variable and test session. The presence of a significant interaction has the implication that learning effects are not equivalent in the specific sub-groups. However, the absence of this interaction would indicate that the demographic variable has no impact on learning. Thus, it is possible that a specific demographic variable could have an impact on overall levels of performance, but improvement across test sessions could remain unaffected. The outcomes of the ANOVAs are presented in Table 5.23.

Table 5.23

*Demographic Effects on Performance of Neuropsychological and Cognitive Tests*

Task/Measure	Gender		Age		Education	
	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction
Snap	$F = 0.36$ $p = .56$	$F = .51$ $p = .77$	$F = 1.20$ $p = .29$	$F = 1.35$ $p = .24$	$F = 0.57$ $p = .45$	$F = 0.77$ $p = .56$
Anagram	$F = 0.06$ $p = .80$	$F = 0.75$ $p = .58$	$F = 3.90$ $p = .06$	$F = 1.41$ $p = .23$	$F = 0.70$ $p = .41$	$F = .039$ $p = .85$
Digit Span						
Forward	$F = 0.17$ $p = .68$	$F = 1.08$ $p = .37$	$F = 1.18$ $p = .28$	$F = 0.54$ $p = .74$	$F = 0.76$ $p = .39$	$F = 0.51$ $p = .76$
Backward	$F = 0.17$ $p = .68$	$F = 0.72$ $p = .60$	$F = 1.64$ $p = .21$	$F = 0.52$ $p = .75$	$F = 3.43$ $p = .07$	$F = 0.95$ $p = .45$
Rey Tangled Lines	$F = 0.01$ $p = .94$	$F = 1.52$ $p = .18$	$F = 3.35$ $p = .08$	$F = 1.40$ $p = .22$	$F = 3.36$ $p = .08$	$F = 0.87$ $p = .50$
Stroop						
Read	$F = 0.69$ $p = .41$	$F = 0.50$ $p = .77$	<b><math>F = 7.38</math></b> <b><math>p = .010</math></b>	$F = 1.07$ $p = .38$	$F = 1.28$ $p = .26$	$F = 0.35$ $p = .87$
Colour	$F = 0.02$ $p = .87$	$F = 0.31$ $p = .90$	<b><math>F = 7.03</math></b> <b><math>p = .014</math></b>	$F = 1.36$ $p = .24$	$F = 1.17$ $p = .28$	$F = 0.20$ $p = .95$
Colour-Word	$F = 0.06$ $p = .80$	$F = 0.75$ $p = .58$	<b><math>F = 16.5</math></b> <b><math>p = .000</math></b>	$F = 1.59$ $p = .17$	<b><math>F = 4.68</math></b> <b><math>p = .04</math></b>	$F = 0.75$ $p = .58$
Phonemic Fluency						
First Responses	$F = 0.35$ $p = .56$	$F = 1.00$ $p = .42$	$F = 1.02$ $p = .32$	$F = 0.41$ $p = .83$	$F = 0.27$ $p = .60$	$F = 2.25$ $p = .06$
Later Responses	$F = 0.01$ $p = .94$	$F = 0.41$ $p = .84$	$F = 1.27$ $p = .27$	$F = 1.64$ $p = .15$	$F = 0.02$ $p = .87$	$F = 0.78$ $p = .56$
Total	$F = 0.05$ $p = .81$	$F = 0.34$ $p = .88$	$F = 1.34$ $p = .25$	$F = 1.63$ $p = .15$	$F = 0.01$ $p = .96$	$F = .40$ $p = .84$
Semantic Fluency						
First Responses	$F = 1.57$ $p = .22$	$F = 1.24$ $p = .29$	<b><math>F = 5.15</math></b> <b><math>p = .03</math></b>	$F = 1.91$ $p = .09$	$F = 0.58$ $p = .45$	$F = 2.28$ $p = .06$
Later Responses	$F = 0.11$ $p = .74$	$F = 1.36$ $p = .24$	<b><math>F = 11.8</math></b> <b><math>p = .002</math></b>	$F = 1.60$ $p = .16$	$F = 3.51$ $p = .07$	$F = 0.75$ $p = .58$
Total	$F = 0.37$ $p = .55$	$F = 2.18$ $p = .06$	<b><math>F = 14.3</math></b> <b><math>p = .001</math></b>	$F = 1.21$ $p = .30$	$F = 4.06$ $p = .06$	$F = 1.98$ $p = .08$

The ANOVA exploring gender effects (17 females and 10 males) established that there were no gender effects on any of the primary measures. Likewise, there were no significant interactions between gender and test sessions. Consequently, gender does not appear to be an issue in evaluating performance on the tasks in this battery.

In order to explore age effects, the sample was divided into those under the age of 70 (N = 17) and those over the age of 70 (N = 10). There were two tasks where age did have an impact on performance. On all three components of the Stroop task, the older participants did not perform as well as those under 70 years of age. The same was true on all three measures of the semantic fluency task, with the older participants producing fewer category exemplars than the younger group. Importantly, age did not interact with test sessions on any of the measures. Learning was equivalent on all tasks for each age cohort.

The distribution of education levels in the normative sample were two people left school before completing high-school, 15 completed high-school, and 10 had some form of university training. The ANOVA exploring education effects involved the 17 people whose formal education did not involve university training versus the 10 people who did. There was only one significant main effect with the less educated participants performing better than their more educated counterparts on the Stroop Colour-Word test. No significant interactions involving test sessions. With one exception, education levels had very little impact on any aspect of performance on the tasks in the battery.

In sum, the demographic variables have no influence on the degree of learning going on in any of the tasks. This finding suggests that the normative sample is appropriate to evaluate any improvements, or lack thereof, in patients with stroke in the same age bands. However, age does become a factor in determining whether an impairment exists on the Stroop and semantic fluency tasks.

## 5.8 Chapter Summary

Table 5.24

*Summary of Task Characteristics*

<b>Task</b>	Linear Improvement	When Significant	Slope & Intercept Independent	Reliability	Benchmark
<b>Snap</b>	N/A	N/A	N/A	Good	Dual Task
<b>Anagram</b>	N/A	N/A	N/A	Good	Dual Task Attenuation of letter length Fwd>Back
<b>Digit Span</b>					
<i>Forward</i>	Yes	S5	No	Acceptable	
<i>Backward</i>	Yes	S3	No	Poor	
					Read>Colo ur>Colour- Word
<b>Stroop</b>					
<i>Read</i>	Yes	S2	No	Good	
<i>Colour</i>	Yes	S2	Yes	Good	
<i>Colour-Word</i>	Yes	S2	Yes	Good	
					Fast start slow finish
<b>Verbal Fluency</b>					
<i>1<sup>st</sup> Response</i>	Yes	S3	Yes	Good	
<i>Later Responses</i>	Yes	S2	Yes	Good	
<i>Total</i>	Yes	S2	Yes	Good	
<b>Semantic Fluency</b>					Fast start slow finish
<i>1<sup>st</sup> Response</i>	Yes	S3	Yes	Acceptable	
<i>Later Responses</i>	Yes	S3	Yes	Good	
<i>Total</i>	Yes	S3	Yes	Good	
<b>Rey Tangled Lines</b>	Yes	S5	Yes	Good	

Table 5.24 summarises the characteristics of the test battery when administered on six occasions over a two-week period to a group of healthy older participants. There is nothing in the data that would suggest any abnormalities in the way in which each task was performed as benchmark effects were replicated in the experiment. More importantly, most tasks showed acceptable psychometric properties. There was also converging evidence of linear learning across the testing sessions and where this did not happen it was due to the introduction of dual task manipulations.

Because meta-analyses have shown that practice effects are not equivalent across tasks (Calamia et al., 2012), learning effects for each task were graphically compared. In Figure 5.14 raw scores on each task have been converted to scaled scores with initial performance being set at a scaled score of 10. Scaled scores for the following sessions were calculated based on initial means and standard deviations, with increases in scaled scores reflecting the degree of improvement across tasks. The Figure indicates that while practice effects were present in all tests, these effects were not as pronounced for forward digit span and Rey Tangled Lines as for on all other tasks.

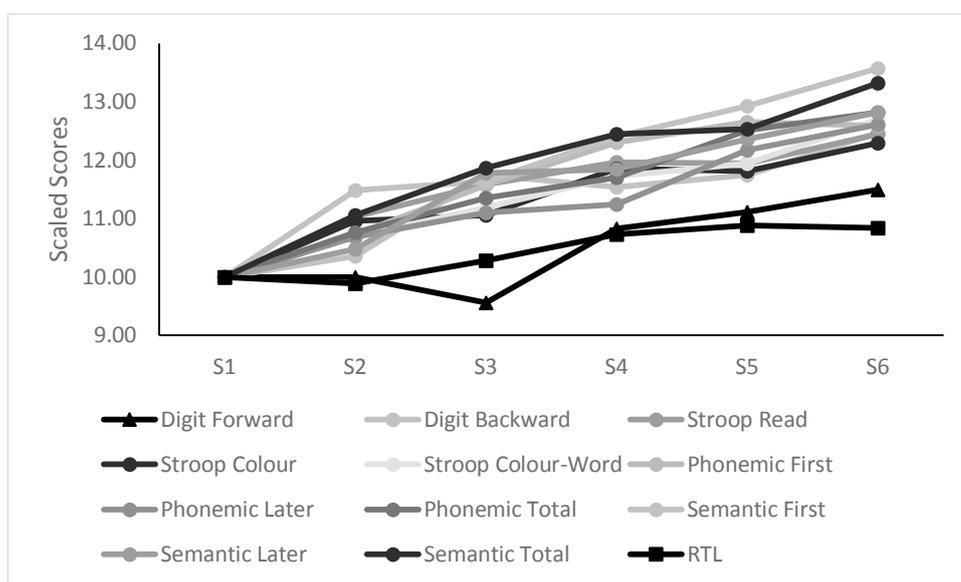


Figure 5.14. Learning effects expressed as scaled scores as function of test measure and test session.

Given the above outcomes there are caveats concerning some of the measures. With forward span, there was comparatively weak learning going on across repeated sessions, and the slope data would suggest that those who learned most were those who were poorest at the first session. Moreover, the reliability of the test was only marginal meaning that learning effects have to be large to warrant consideration. The same was true of backward span. Although learning was much the same as in other tasks, again the rate of learning was related to initial levels of performance and reliability was again low. Consequently, the digit span data for the patients with stroke should be treated with some caution. For the Rey Tangled

Lines, -practice effects were small and only emerged in the fifth session. This may have ramifications for the ability to detect performance among patients with stroke. For the other tasks, however, one can use the current data as a source of normative data to compare absolute levels of performance and learning effects in those who are recovering from stroke.

## **Chapter 6: Introduction to Case Studies**

### **6.1 Introductory Issues**

In the following two chapters, five case studies are described where patients who have suffered cerebrovascular trauma were administered the test battery. Each of the case studies addresses two basic issues: what aspects of cognition have been impaired by a stroke, and how behaviour changes across test sessions. In Chapter 7 these issues are addressed in the context of the neuropsychological tests, and in Chapters 8 they are addressed in the context of the cognitive tests. Logically, the outcomes of each case study can be defined in terms of four different profiles. Firstly, the most positive outcome is that there is improvement on all tasks. Secondly, it is possible that there will be improvement on the unimpaired tasks but no improvement on the impaired tasks. Thirdly, it is possible that there might be no improvement on unimpaired tasks but improvement on the impaired tasks. The final outcome is where there is no improvement on any of the tasks; an outcome that is the least positive of the four profiles.

#### **6.1.1 Heterogeneity**

The organisation and presentation of the case studies are influenced by a number of factors. As indicated earlier in Chapter 2, the profiles of impairments among patients with stroke are heterogeneous, and it is unlikely that any two case studies will show exactly the same profile of impairments. With different profiles of impairment, there is no a priori reason to present one case study before another.

#### **6.1.2 Comorbidities**

The presence of comorbidities is also an issue. The prevalence data reported by the Australian Institute of Health and Welfare (Australian Institute of Health and Welfare, 2013) showed that 87% of patients with stroke are 55 years or older, with 45% of patients with stroke being the 75 years and older. Almost identical rates were found in a large Scottish sample (Gallacher et al., 2014). In all likelihood a stroke patient is going to be in an age

bracket where some cognitive decline has already occurred through normal ageing (Anstey & Low, 2004). Moreover, stroke as a disease is associated with high levels of comorbidity. For example, Gallacher et al. (2014) explored the prevalence of 39 comorbidities among 35,690 Scottish people who had suffered a stroke, and compared these prevalence rates with a control group of 1,388,688 controls. Of the stroke group, only 5.8% did not have any comorbid condition. If one considered the 75+ age group, then only 2.5% patients had no comorbidity. This study also explored the number of comorbid conditions each patient reported. In the 75+ age bracket, 40% had between one to three comorbid conditions, 40% had between four and six comorbid conditions and 13% had seven or more comorbid conditions. Caughey et al. (2010) reported similar prevalence data in 2,087 patients with stroke aged 65 years and older in an Australian sample. They reported that only 12% of this sample did not have a comorbid chronic disease and 64% had multimorbidities. Given these data sets, it would appear that the prototypical stroke patient will be over 75 years old and have multimorbidities. The profile of multimorbidities is likely to vary from patient to patient. Therefore, not only are the effects of stroke quite heterogeneous, but also the profile of comorbid conditions is likely to be equally heterogeneous.

In these studies it is clear that not all the comorbid conditions are likely to have a direct impact on cognitive performance (e.g., constipation). However, conditions like Parkinson's disease, substance abuse and depression are all known to have a detrimental impact on cognitive performance (Bohnen et al., 2014; Bott et al., 2014; Pfeiffer, Løkkegaard, Zoetmulder, Friberg, & Werdelin, 2014). In the Gallacher et al. (2014) study these conditions were much more likely to be observed in the patients with stroke than in the control group (Odds Ratio from 1.2 to 2.0). Moreover, the Gallacher et al. study also showed a highly significant correlation between the number of comorbidities a person had and the number of medications they took. Thus, the patients with stroke were characterised by a greater range of comorbidities and a larger cocktail of medications than the control sample.

The heterogeneity of impairment and multimorbidity profiles, and issues with polypharmacy, all converge on the fact that it is virtually impossible to control for any of these three factors. Each stroke patient is likely to have their unique impairments, his or her unique comorbidities and unique combination of drugs. Moreover, many of the comorbidities may be unknown. Even if comorbidities are mentioned on patient charts, the accuracy of the diagnosis is not assured, and the actual impact on cognitive performance is not known. The one positive aspect of multimorbidities and polypharmacy is that these factors should either inhibit learning or produce decrements in performance. None of them should result in improvements in performance.

While it is not possible to identify all relevant comorbidities, the approach taken in presenting the case studies is to first present those cases in which there are no known comorbidities. These cases are followed by those in which possible comorbidities have been identified.

### **6.1.3 Normative Sample versus Matched Controls**

The third issue involves comparison groups. In Chapter 5, the utility of normative samples versus matched controls was discussed, and it was noted that there were advantages and disadvantages of both approaches to comparing performance with clinical samples. Chapter 5 also showed that with regards to learning outcomes (changes across sessions), demographic factors such as gender, age and education had no impact. The implication is that the normative sample can also serve the same function as a control group when improvement across sessions is examined. However, while there were no gender and no consistent education effects on absolute levels of performance, there were significant age effects on two of the tasks. Such differences suggest that an age-matched sample might be appropriate for these two tasks. When reporting the case studies, performance of each stroke patient is compared to the normative sample described in Chapter 5, and to five participants from the normative sample who were in the same age bracket as the patients with stroke. Four of the

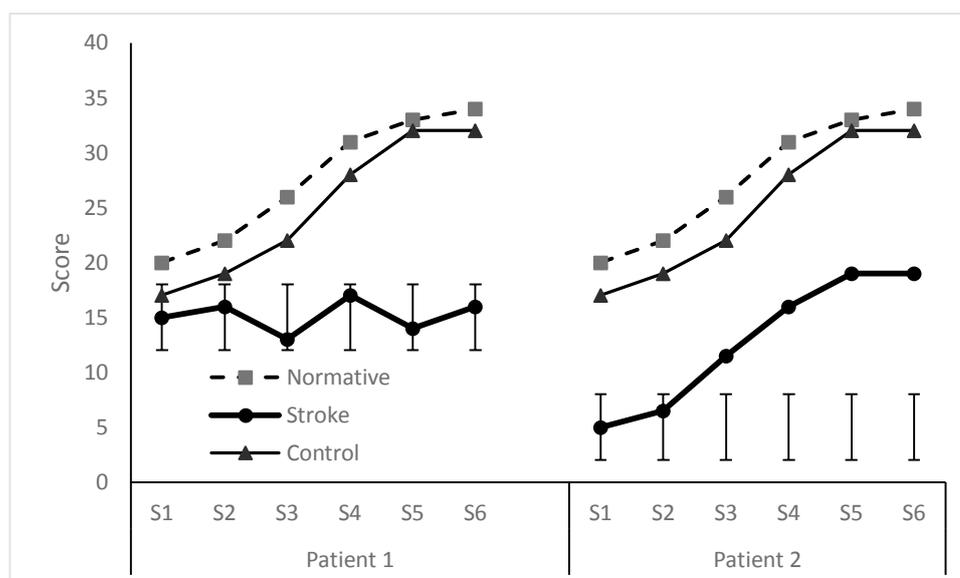
patients with stroke were between 79 and 82 years old. The five controls were aged between 79 and 87. The fifth stroke patient was 55 years of age, and the five controls for this patient were aged between 55 and 60.

#### **6.1.4 Evaluating Learning**

The previous experiments addressed the issue of learning using two procedures. The first was via analysis of variance techniques, and the second involved temporal stability measures. Using particular error terms and specific cut-off points enabled temporal stability confidence intervals to be constructed, which gave identical results to those generated by the analyses of variance. These two procedures were used in anticipation of the case studies. In the case studies, analyses of variance cannot be conducted where an individual is involved, and so alternative measures of learning need to be adopted. While there have been a number of methods generated for measuring cognitive change (Frerichs & Tuokko, 2005), several are based on a single re-test, and it is not clear how they could be used to assess change with multiple administrations. Several indices have been used to address changes at the individual task level, and a slope methodology has been used to evaluate more global changes where performance is collapsed across tasks.

**6.1.4.1 Individual Tasks.** The first measure of leaning is presented by converting each participant's raw score into z-scores for each session. These data are presented in tabular form. While the z-scores are useful, they are not used as the primary source of evaluating learning.

Changes in performance across testing sessions are presented graphically using two formats. In the first instance, raw scores on each task are presented for the stroke patient, the average of the matched control groups, and the normative sample using the format depicted in Figure 6.1.



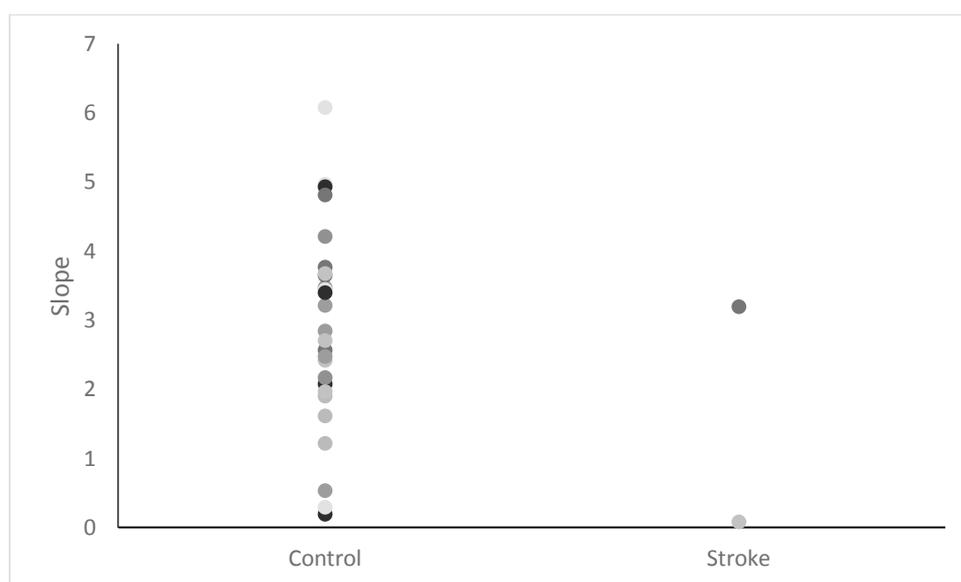
*Figure 6.1.* Performance of two hypothetical patients with stroke on a cognitive task as a function of test session.

Figure 6.1 provides a graphical account of changes in performance across testing sessions for the two hypothetical patients with stroke. It is apparent that Patient 1 does not improve much across test sessions. The normative sample, the matched controls and the stroke patient improve with practice, and improve at the same rate.

In Chapter 5, learning outcomes in the normative sample were evaluated by constructing 95% test confidence intervals around the baseline mean. Learning was deemed to have occurred if performance emerged outside the confidence band on subsequent sessions. Learning in the patients with stroke can be evaluated by applying the test confidence band derived from the normative sample to the stroke patient's performance. Again, learning is judged to have occurred if performance emerges outside the confidence band. Figure 6.1 confirms that Patient 1 shows no learning across the six sessions, whereas Patient 2 has improved above baseline by Session 3.

The second format for evaluating learning involves slope measures derived from the best fitting regression equations. Assessment of slope values will be done both graphically and in tabular form for all primary measures except for Span and Anagram tasks where slope values are compromised by dual task effects. In Figure 6.1 the slope for Patient 1 is .09, for

Patient 2 it is 3.20, and for the normative sample it is 3.08. In Figure 6.2 the slope estimates have been calculated and plotted for each member in a hypothetical normative group, along with the slope values of the two patients with stroke. The slope values of the patients with stroke relative to the normative sample are readily apparent, in that Patient 1 shows little evidence of learning, the value is low, and more importantly, the slope is at the edge of the normative sample distribution. In contrast, Patient 2 has a substantial slope estimate and this estimate is near the mean of the slopes generated by the normative sample.



*Figure 6.2.* Slope values for control and patients with stroke on a hypothetical cognitive task.

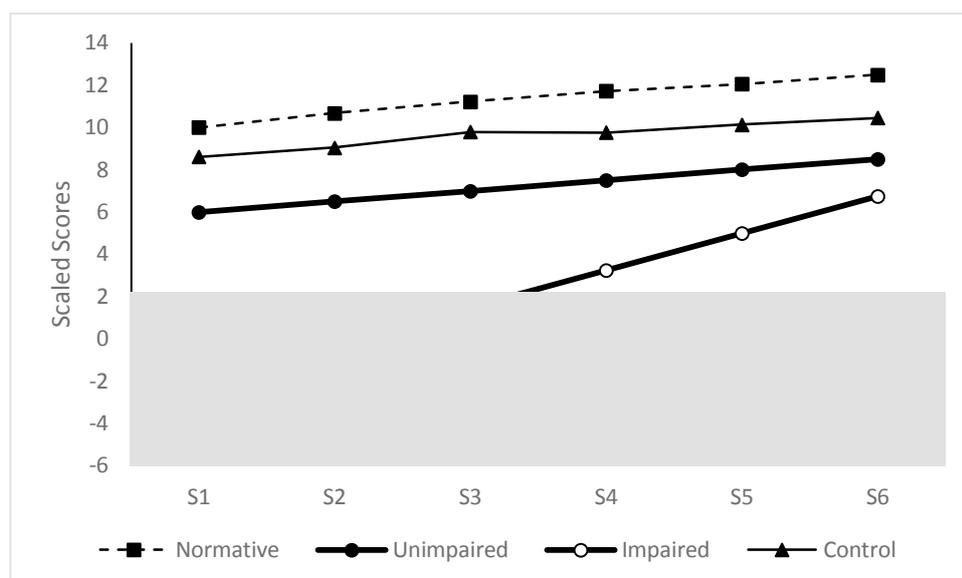
As was the case with mean performance, confidence intervals for the slopes can be derived based on normative performance and z-scores can be calculated for the patients with stroke, in this case  $z = -1.88$  for Patient 1 and  $z = .09$  for Patient 2. The z-scores can also be transformed into percentile rankings, such that only 3 % of slopes in the population should be below that of Patient 1, but 54% of slopes will be less than that showed by Patient 2. Thus, in tabular form the mean slope and 95% confidence intervals based on the normative sample are reported. The slope of the stroke patient, the z-score and percentile rankings associated with that slope are also reported. However, it should still be kept in mind that values close to zero indicate that no learning is occurring, irrespective of z-score or percentile rankings.

One additional form of learning is also evaluated but in this instance it is specific to the Anagram task and is reported in Chapter 8. In the normative sample, there were robust letter-length effects in the early sessions with solution times for 5-letter anagrams being appreciably slower than for 6- and 7-letter anagrams. However, in the latter sessions, performance on the 5-letter anagrams improved markedly such that the length effects were attenuated in the later sessions. The attenuation of letter-length effects in the patients with stroke will also serve as a marker for learning in this task.

**6.1.4.2 Comparing initially impaired and unimpaired tasks.** While performance at the individual task level is important, useful diagnostic information is typically derived when performance has been collapsed across tasks (Darby et al., 2002). Moreover, from a theoretical perspective it has been assumed that on unimpaired tasks, patients with stroke will show normal practice effects. The strength of learning should be the same as for the normative sample. In the case of the impaired tasks, the expectation was that if recovery were present, this would be evident in a steeper learning curve. To evaluate the learning involved in impaired tasks and in unimpaired tasks, the raw scores need to be converted into a common metric. The Nordvik et al., (2012) method of graphically summarising changes in terms of scaled scores are adopted and adapted (see Figure 2.3). In the first instance the z-scores for each task are converted to scaled scores that have a mean of 10, a standard deviation of 3. Scaled scores are usually restricted to a range of between 1 and 19, although no such constraint is applied in the current summaries. The scaled scores are then averaged across tasks. In the case of the normative sample and the control group the average is based on all tasks. With the patients with stroke, separate averages were calculated for initially impaired performance and initially intact measures. Performance of the normative group on the first test session was set to a scaled score of 10. The initial score of the normative sample is used as the reference point for all subsequent comparisons and improvement is represented by increases in scaled scores across testing sessions.

Figure 6.3 presents the summary of the normative sample, a matched control group, and a hypothetical stroke patient. Here it is apparent that learning rates for the normative sample, the control group and for the unimpaired domains of the patients with stroke all improve with repeated testing and improve at the same rate. A best case recovery scenario is presented for initially impaired performance where there is a rapid improvement across sessions. The grey area in the Figure represents scaled scores in the impaired range. This corresponds to a scaled score of 4.12, which is the equivalent of a z-score of -1.96. In the case of the hypothetical stroke patient, performance on the initially impaired tasks, improve rapidly and performance is within normal range functioning by Session 4.

Slopes of these curves can be constructed in the same way as for the individual tasks and 95% confidence intervals generated to determine the extent to which the performance of the stroke patient differs to that of the normative group. In Figure 6.3 it would be expected that the slopes of the control group and the unimpaired tasks of the patients with stroke would fall within the 95% confidence band. If this were the case then one could conclude that the stroke patient is showing normal learning on the unimpaired tasks. For the impaired tasks, it would appear that the slope value is outside the upper limit of 95% confidence interval, which would be indicative of recovery taking place. Likewise, although not depicted in Figure 6.3, a slope value that was below the lower 95% confidence limit would be indicative of little or no recovery.



*Figure 6.3.* Overall summary of stroke patient's performance expressed in averaged scaled scores.

## 6.2 Case Study Format

The case studies address these issues using a standard format. In Chapter 7 the format involves background information about the stroke patient, description of the testing methodology, the obtained results for the neuropsychological tests, and an overall evaluation of the stroke's impact on the participant's cognitive functioning. In Chapter 8, the results of the cognitive tests are presented and again the impact and implications of the outcomes are described.

### 6.2.1 Background

The case studies begin with an introduction that provides some background information about the stroke patient, although not as much as would normally be presented when attempting to explain why performance varies as it does, why some patients with stroke have improved and others not, or which particular cognitive domain has been affected. As argued before, the battery does not provide enough depth or breadth to address those issues.

Basic demographic information is reported, followed by a description of the precipitating event and the eventual diagnosis made by the medical staff. Premorbid estimates

of cognitive functioning are provided and cognitive screening outcomes are reported when available. Any known pre-existing health factors or known comorbidities are also reported.

### **6.2.2 Method**

The methodological issues associated administering the test regime are then described. The tests administered in each of the test sessions are presented in tabular format. The tests were always administered in the same order: phonemic fluency, semantic fluency, Anagram task, Digit Span, Stroop task, Rey Tangled Lines, and Snap. Information is provided regarding when the first testing session was held relative to when the stroke occurred, the length of each testing session and the duration between successive testing sessions. Testing notes concerning environmental and participant factors are reported to provide the context in which the tests were administered. Factors such as fatigue, arousal, and ambient noise could all impact on performance resulting in across-session variability. Such information might be useful for interpreting the obtained patterns of performance.

### **6.2.3 Results**

The results of the testing regime are described in terms of absolute levels of performance and learning effects. Absolute levels of the stroke patient's performance are evaluated against performance of the normative sample. Each stroke patient's performance on each measure is represented as a z-score based on the mean and standard deviation of the normative sample. Z-scores of greater than -1.96 are considered to be in the clinical range. Z-scores across all measures for all testing sessions are presented in tabular form. Raw scores on each task are then presented in graphical form where learning effects can be directly compared. These Figures also present test temporal stability confidence intervals around the stroke patient's performance to provide the first statistically derived measure of learning. The regression slope data, which serve as the second index of learning, are then presented in tabular and graphical form. Finally, performance collapsed across intact and impaired tasks is presented based on scaled scores. These slopes are evaluated using 95% confidence intervals.

Once the primary aspects of impairment and learning are assessed, the results of the individual tasks will be examined. The intent here is to describe any additional points that are unique to each task or have not been addressed in dealing with the primary objectives of the results. One example could involve the differences between early and later responses on the verbal fluency tasks.

#### **6.2.4 Discussion and Overall Evaluation**

The discussion and overall evaluation provides an assessment of impairment and learning both at the task level and at the composite level. The aim here is to make a judgement concerning the extent to which practice effects and/or recovery have been detectable in the data.

## Chapter 7 Case Studies: Neuropsychological Tasks

### 7.1 Introduction

This chapter introduces five case studies of patients who were admitted to the stroke rehabilitation ward of a large metropolitan hospital in Brisbane, Australia. All had been admitted because they had suffered cerebrovascular trauma. In the first three case studies, the patients (GL, MF and KS) have no previously identified comorbid conditions that have well-documented, detrimental effects on cognitive performance. The remaining two (LL and TB) have diagnoses of Parkinson's disease. TB also experienced visual impairment issues associated with the stroke which precluded the administration of some tests in the battery. Given the advanced age of some of these patients it is highly probable that multimorbidities are present, although not identified.

### 7.2 Case Study 1: GL

#### 7.2.1. Participant Background

GL is an 82-year-old male, retired draughtsman who was admitted to hospital with a subdural hematoma.

**7.2.1.1 Precipitating event.** GL over balanced at home and fell off a chair. It was reported that his head made soft contact with the wall but he did not experience any pain or loss of consciousness. Hours later, GL experienced an acute onset of upper limb weakness, slurring of his speech and difficulty moving.

**7.2.1.2 Diagnosis.** Subsequent medical investigation revealed that GL had suffered a subdural haematoma causing damage to both hemispheres of his brain. He was referred to the rehabilitation ward for physical rehabilitation of his motor function.

**7.2.1.3 Cognitive performance.** GL's premorbid intellectual functioning as assessed by WAIS-III sex, age and education norms was evaluated to be FSIQ = 107 [90% CI 84, 130]. On admission to hospital, the MMSE was administered and GL scored 29/30 (normal cognition).

**7.2.1.4 Other factors.** GL's medical history includes laryngeal cancer and coronary artery bypass surgery.

## 7.3 Method

### 7.3.1 Test Administered

The tests were administered to GL on six occasions before he was discharged from the rehabilitation ward. The tests that were administered on each session are summarised in Table 7.1. The tests were always administered in the same order: Rey Tangled Lines, phonemic fluency, semantic fluency, Anagram task, Digit Span, Stroop, and Snap task. During the initial test session, standardised instructions were read and the tasks were administered in accordance with the procedure described in Chapter 3. On subsequent tests, abbreviated instructions to each test were administered if it was clear that the patient had remembered the tasks from previous sessions.

Table 7.1

*Tests Administered to GL across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<b>Digit Span</b>							
<i>Forward</i>	Single	✓	✓	✓	✓	✓	✓
<i>Backward</i>	Single	✓	✓	✓	✓	✓	✓
<b>Stroop</b>							
<i>Read</i>	Single	✓	✓	✓	✓	✓	✓
<i>Colour</i>	Single	✓	✓	✓	✓	✓	✓
<i>Colour-word</i>	Single	✓	✓	✓	✓	✓	✓
<b>Verbal Fluency</b>							
<i>Phonemic</i>	Single	✓	✓	✓	✓	✓	✓
<i>Semantic</i>	Single	✓	✓	✓	✓	✓	✓
<b>Rey Tangled Lines. (RTL)</b>	Single	✓	✓	✓	✓	✓	✓

### 7.3.2 Test Sessions

The first test session was held 5 days after GL suffered his stroke. He was administered the test battery on six occasions during the following 20 days until he was discharged from the rehabilitation ward. Table 7.2 presents information about the duration of each test session and the interval between successive testing sessions.

Table 7.2

*Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions*

	Testing Session					
Timing	S1	S2	S3	S4	S5	S6
Duration	50	50	55	45	45	45
Interval	-	2	2	7	4	3

### 7.3.3 Test Notes

Session 1: GL provided written and verbal consent to participate in the study. Testing took place at his bedside. Instructions were explained and GL was able to complete the Rey Tangled Lines and phonemic fluency. Semantic fluency was administered, however GL was unable to stay awake and the session was terminated. On the following day GL was seen immediately after his physiotherapy session and he presented as energetic and engaged. He was able to complete the remainder of the test battery. As GL had fallen asleep during the semantic fluency task on the previous day, it was administered for a second time in order to obtain valid data on this measure.

Session 2: GL performed well in session two, remembering the assessor and previous instructions for the tasks. The complete test battery was administered, with a single interruption during Snap which caused GL to lose concentration briefly. During this period he missed two snaps, but regained concentration shortly after.

Session 3: GL reported being tired during this session, but was able to complete the battery without difficulties.

Session 4: GL was seen immediately after his physiotherapy session. He presented as energetic and engaged and was able to recall all instructions. Despite this, he performed poorly during this session. On further investigation, GL reported difficulty concentrating. The single event version of the dual tasks was administered for the first time in the Snap and Anagram tasks

Session 5: GL performed well during this session, with consistent concentration and knowledge of the required tasks. The two event dual task condition was introduced in this session.

Session 6 & Feedback session: All tests were again administered without interruption. At the conclusion of the tests preliminary feedback was given to GL on his performance over the six sessions.

## **7.4 Results**

### **7.4.1 Absolute Performance**

**7.4.1.1 Initial levels of impairment.** GL's absolute levels of performance across the five tests are presented in Table 7.3 and Figure 7.1. At the initial testing, GL's performance was outside the 95% confidence levels on seven of the twelve primary measures. The specific impairments were in the read and colour naming components of the Stroop task, the initial responses in the phonemic fluency task, all three measures in the semantic fluency task and in completion times on the Rey Tangled Lines test. There were no impairments in digit span, colour-word naming scores in the Stroop task, and later responses and total scores in the phonemic fluency task.

Table 7.3

*GL's Performance Across Tasks Expressed in Z-scores*

Task	Test Sessions					
	1	2	3	4	5	6
<b>Digit Span</b>						
Forward	-0.28	-2.02	-1.06	-1.36	-3.29	-1.83
Backward	0.10	-2.02	-3.00	-1.68	-1.41	-1.72
<b>Stroop</b>						
Read	<b>-3.92</b>	-4.95	-4.43	-3.72	-3.74	-4.90
Colour	<b>-3.45</b>	-2.79	-2.64	-2.64	-3.09	-2.69
Colour Word	-1.80	-2.13	-2.24	-1.42	-2.07	-2.01
<b>Phonemic Fluency</b>						
1 <sup>st</sup> Responses	<b>-2.56</b>	-1.15	-2.05	-2.12	-2.03	-2.35
Later Responses	-1.40	-1.57	-2.24	-2.09	-1.87	-1.97
Total	-1.90	-1.60	-2.39	-2.37	-2.02	-2.30
<b>Semantic Fluency</b>						
1 <sup>st</sup> Responses	<b>-3.65</b>	-1.81	-2.83	-2.21	-2.21	-2.93
Later Responses	<b>-2.57</b>	-2.10	-3.76	-3.63	-3.04	-2.98
Total	<b>-2.75</b>	-2.03	-3.75	-3.46	-2.80	-3.07
Rey Tangled Lines	<b>-2.76</b>	-7.32	-12.13	-4.79	-5.84	-4.10

Note: Numbers in bold font indicate impaired performance on initial test

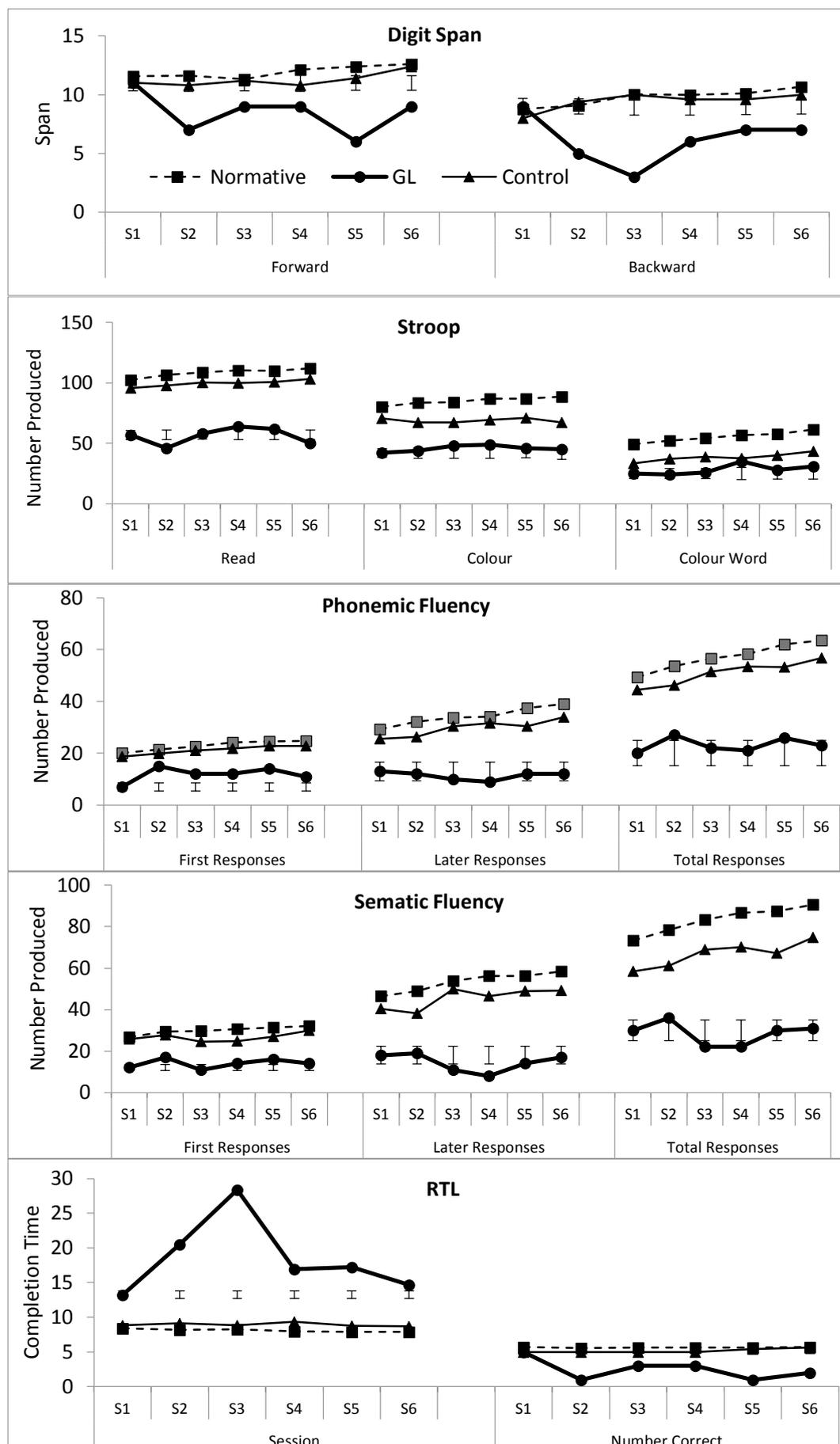


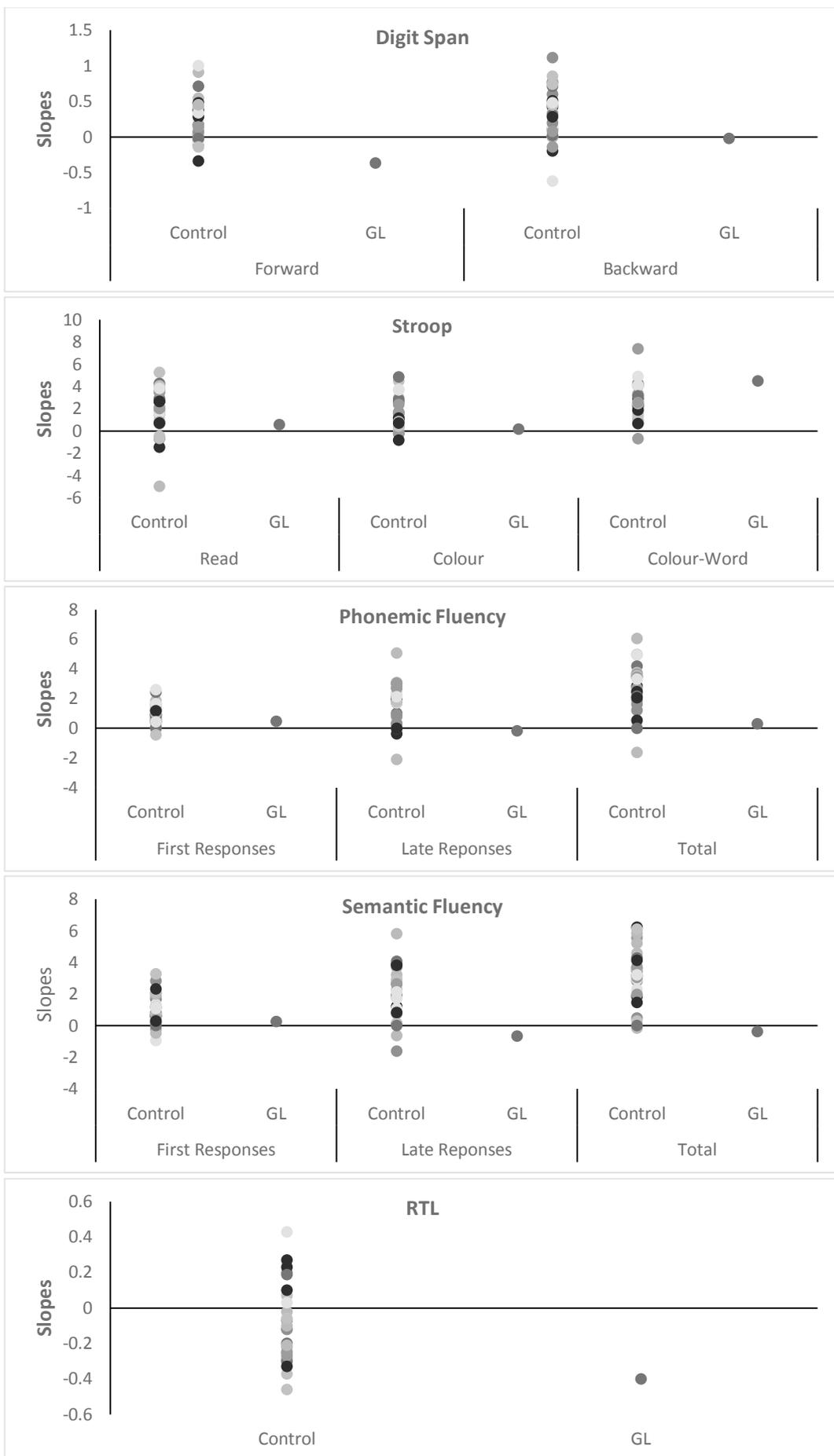
Figure 7.1. GL's performance relative to normative sample.

## 7.4.2 Learning

**7.4.2.1 Temporal stability confidence intervals.** Temporal stability confidence intervals are presented around GL's scores in Figure 7.1. In all tasks, with the exception of first responses in the phonemic fluency tasks, there was little systematic improvement across tasks. Performance improved or deteriorated outside the confidence band, on occasion, but there did not seem to be any systematic and sustained improvement across sessions. In the case of first responses in the phonemic fluency task, there was a significant improvement above baseline that was maintained across subsequent sessions. However, if one looks at the changes that occurred across Sessions 2 to 6, there is little evidence that there is continuing improvement across sessions.

**7.4.2.2 Slopes.** The slopes of the best-fitting regression lines for both GL and the normative sample are presented in Figure 7.2 and Table 7.4. Figure 7.2 confirmed the absence of any marked change across sessions as in most tasks GL's slopes were very close to zero. Moreover, Table 7.4 indicates that on all tasks, save the Rey Tangled Lines, at least 70% percent of the normative sample was showing superior learning rates compared to GL. Thus, across tasks there is very little evidence supporting increased learning across trials.

There were three tasks where GL's slopes were outside the 95% confidence intervals: forward digit span, late responses and total scores on the semantic fluency task, and the Rey Tangled Lines test. In the case of forward digit span, his performance deteriorated across sessions. In the case of the semantic fluency task, GL's slope values were similar to those observed on other tasks. What differs here from other measures is the degree of improvement in the normative sample relative to other tasks. In the case of the Rey Tangled Lines test, the slope would suggest that there was very strong learning in this task. However, Figure 7.1 indicates that GL's performance on this task was extremely variable and that the high slope value was an artefact of this variability.



*Figure 7.2.* GL's slope values for all tasks compared to the normative sample.

Table 7.4

*GL's Learning Slopes Relative to the Normative Sample*

Task	Control	GL	Z	%ile
<b>Digit Span</b>				
Forward	.22 (-.42)	-0.37	-1.83	3
Backward	.28 (-.88)	-0.02	-0.51	30
<b>Stroop</b>				
Word	1.68 (-2.06)	0.54	-0.52	30
Colour	1.59 (-1.41)	0.62	-0.63	26
Colour Word	2.23 (-.35)	1.45	-0.59	28
<b>Phonologic Fluency</b>				
1st Responses	.97 (-.41)	0.48	-0.70	24
Later Responses	1.74 (-1.19)	-0.17	-1.27	10
Total	2.71 (-.55)	0.31	-1.44	18
<b>Semantic Fluency</b>				
1 <sup>st</sup> Responses	1.06 (-.65)	0.28	-0.89	19
Later Responses	2.13 (-1.14)	-0.65	-1.66	5
Total	3.31 (-.26)	-0.37	-2.02	2
Rey Tangled Lines	-.08 (.35)	-0.40	1.46	93

Note: Numbers in parentheses in Control column represent the lower limit of 95% confidence interval.

### 7.4.3 Impaired and Unimpaired Tasks

The outcomes of GL's performance are summarised in Table 7.5 where the tasks have been organised in terms of whether or not impairment was detected at the initial test session. Tasks that were not in the clinical range are presented first and those in the clinical range are

presented afterwards. The degree of learning that occurred across sessions is evaluated using a Yes/No judgement for both test confidence interval and slope measures of learning.

Table 7.5

*Summary of Performance Outcomes for GL*

Task	Impairment	Learning	
		Conf. Intervals	Slopes
Digit Forward	N	N	N
Digit Backwards	N	N	N
Stroop – ColourWord	N	Y	Y
Phonemic - Later	N	N	N
Phonemic - Total	N	N	N
Stroop - Read	Y	N	N
Stroop - Colour	Y	N	N
Phonemic - First	Y	Y	?
Semantic - First	Y	N	N
Semantic - Later	Y	N	N
Semantic - Total	Y	N	N
Rey Tangled Lines	Y	N	N

Formal assessment of the learning for impaired and unimpaired tasks is summarised in Figure 7.3 where average scaled scores are presented. GL's performance on both initially impaired and initially unimpaired task look very similar, with neither showing any signs of regular and consistent learning. The regression slope for the normative sample was .48 [95% CI = .41, .57]. GL's slope for the unimpaired tasks was -.07 (SE = .03) and for the impaired tasks was .15 (SE = .04). Both of these were outside the lower limit of the confidence interval defining normal learning. GL does not show learning effects in either impaired or unimpaired tasks.

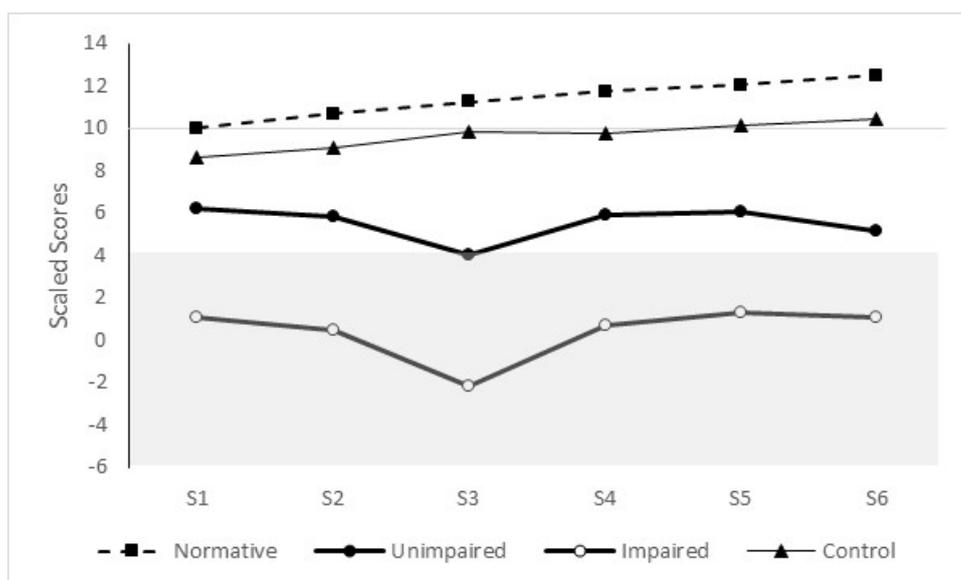


Figure 7.3. GL's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.

### 7.5 Discussion

At initial testing GL was within normal range functioning in 7 of the 12 primary measures. In the tasks where GL was initially within normal limits there was very little evidence indicating that performance improved with repeated testing. At the level of the individual tasks, GL was in the normal range on initial test on the two digit span measures, colour-word naming scores in the Stroop task, and later responses and total scores in the phonemic fluency task. Both slope estimates and temporal stability measures indicated that there was no strong and consistent learning in these tasks. This was confirmed when performance across tasks was averaged. Again there was no indication of normal practice effects in the unimpaired tasks.

The same was true of the tasks and measures that were originally in the impaired range. At the task level with both slope and temporal stability measures, there was little evidence for consistent improvement. Likewise, when performance across tasks was collapsed, there was no evidence for learning.

## 7.6 Case Study 2: MF

### 7.6.1 Participant Background

MF is an 82-year-old, female, retired teacher. She was admitted to hospital with a diagnosis of a right frontal lobe subacute infarct.

**7.6.1.1 Precipitating event.** MF was admitted to hospital with a diagnosis of pneumonia. However over a period of 10 days, MF's health deteriorated and she began to show symptoms of lower limb weakness.

**7.6.1.2 Diagnosis.** Medical tests revealed that MF suffered a right frontal lobe subacute infarct in her cingulate cortex. MF suffered no weakness in her upper limbs, but experienced reduced power in her lower limbs.

**7.6.1.3 Cognitive performance.** Premorbid estimates of MF's intellectual functioning indicated a FSIQ = 106 [90% CI = 83, 129]. On admission MF undertook a cognitive assessment as part of her Occupational Therapy assessment. The results revealed that the areas of coordination, sensation, primary visual skills and perception were all intact. In addition, a speech therapy assessment indicated that MF suffered a right hemisphere cognitive problems in naming and abstract reasoning (phonemic fluency= 46, Boston Naming= 12/15 and abstract naming 2/10).

**7.6.1.4 Other factors.** MF had a history of Type 2 diabetes.

## 7.7 Method

### 7.7.1 Test Administered

The tests were administered to MF on five occasions before she was discharged from the rehabilitation ward. The tests that were administered on each session are summarised in Table 7.6. The tests were always administered in the same order: Rey Tangled Lines, Phonemic fluency, semantic fluency, Anagram task, Digit Span, Stroop test, and Snap task. During the initial test session, standardised instructions were read and the tasks were administered in accordance with the procedure described in Chapter 3. In subsequent tests,

abbreviated instructions to each test were administered if it was clear that the patient had remembered the tasks from previous sessions.

Table 7.6

*Tests Administered to MF across Five Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Digit Span</i>							
<i>Forward</i>	Single	✓	✓	✓	✓	✓	
<i>Backward</i>	Single	✓	✓	✓	✓	✓	
<i>Stroop</i>							
<i>Read</i>	Single	✓	✓	✓	✓	✓	
<i>Colour</i>	Single	✓	✓	✓	✓	✓	
<i>Colour-word</i>	Single	✓	✓	✓	✓	✓	
<i>Verbal</i>							
<i>Fluency</i>							
<i>Phonemic</i>	Single	✓	✓	✓	✓	✓	
<i>Semantic</i>	Single	✓	✓	✓	✓	✓	
<i>Rey Tangled</i>	Single	✓	✓	✓	✓	✓	
<i>Lines (RTL)</i>							

### 7.7.2 Test Sessions

The first test session was held 3 days after having suffered the stroke. She was administered the test battery on five occasions during the following 18 days until she was discharged from the rehabilitation ward. Table 7.7 presents information about the duration of each test session and the interval between successive testing sessions.

Table 7.7

*Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions*

Timing	Testing Session					
	S1	S2	S3	S4	S5	S6
Duration	40	35	35	40	35	
Interval	-	4	3	4	5	-

### 7.7.3 Test Notes

MF provided written and verbal consent to participate in the study. Throughout the sessions MF remained interested and alert and she remembered all instructions and performed well on most occasions. She would often ask for feedback and liked to see whether she had improved from last session. The only exception to MF's consistently solid performance was during session five, where she reported being very tired from a previous discussion. She was able to complete the Rey Tangled Lines, Phonemic and semantic fluency tests but performed poorly. The session was terminated and completed a day later. During this make-up session, MF reported being breathless after physio and found it difficult to maintain an appropriate breathing rhythm for rapid speech (Stroop task). Overall, session five was conducted when MF felt fatigued and her performance decreased.

## 7.8 Results

### 7.8.1 Absolute Performance

**7.8.1.1 Initial levels of impairment.** MF's absolute levels of performance across the seven tests are presented in Table 7.8 and Figure 7.4. At the initial testing, MF's performance was outside the 95% confidence levels on 5 of the 12 primary measures. The specific impairments were in the read, colour naming and colour-word components of the Stroop task, and the later and total responses to the semantic fluency task. There were no impairments in digit span, phonemic fluency, first responses in semantic fluency, or the Rey Tangled Lines Test.

Table 7.8

*MF's Performance Across Tasks Expressed in Z-scores*

	Test Sessions					
	1	2	3	4	5	6
<b>Digit Span</b>						
Forward	-1.71	-2.02	-1.06	-0.93	-2.26	-
Backward	0.10	-1.03	-0.87	-0.42	-1.86	-
<b>Stroop</b>						
Read	<b>-2.89</b>	-2.83	-2.68	-3.00	-3.74	-
Colour	<b>-2.55</b>	-1.73	-1.84	-2.16	-2.86	-
Colour Word	<b>-2.02</b>	-1.90	-2.24	-2.00	-2.14	-
<b>Phonemic Fluency</b>						
1 <sup>st</sup> Responses	-0.99	-0.79	-0.71	-0.55	-1.27	-
Later Responses	-0.79	-0.64	-0.73	-0.42	-0.69	-
Total	-0.93	-0.75	-0.80	-0.52	-0.90	-
<b>Semantic Fluency</b>						
1 <sup>st</sup> Responses	-1.69	-1.02	-1.58	-2.21	-2.51	-
Later Responses	<b>-2.32</b>	-1.70	-2.67	-2.50	-2.35	-
Total	<b>-2.06</b>	-1.49	-2.46	-2.64	-2.40	-
Rey Tangled Lines	-1.83	-1.41	-1.33	-1.02	-0.66	-

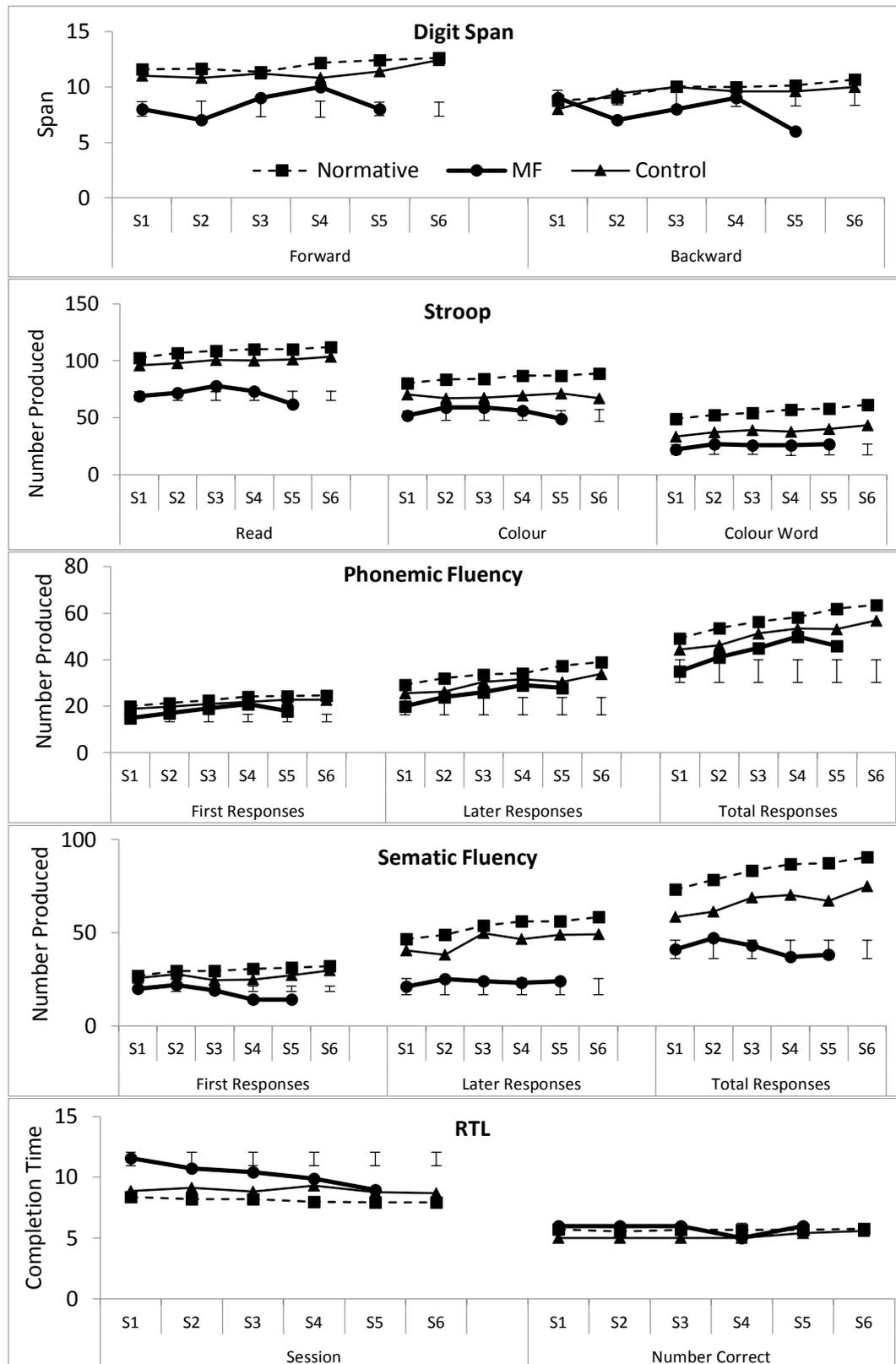


Figure 7.4. MF's performance relative to normative sample.

## 7.8.2 Learning

**7.8.2.1 Temporal stability confidence intervals.** The phonemic tasks and Rey Tangled Lines showed systematic improvements across the task. In the case of phonemic fluency, there was a significant continual improvement above baseline for first responses, later responses, and total responses. The exception to this was on the last session, which saw a minor decrease in performance in both first and later responses. Despite this, MF's scores remained above of the 95% confidence interval from Session 2. There were regular gains on the Rey Tangled Lines Tests as indicated by continuous improvement in completion times. By Session 2 performance had improved to be outside the 95% confidence interval. Correctly identifying the target item remained consistent throughout testing, with the exception of Session 4 where there was a minor decline in identifying the correct targets. In all other tasks there was little systematic improvements across tasks.

**7.8.2.2 Slopes.** MF's learning outcomes are presented in Figure 7.5 and Table 7.9. It is clear that uniform learning was not present in all tasks. There were positive learning outcomes in the Rey Tangled Lines, forward digit span and in the phonemic fluency tasks. In each of these tasks MF's slope value was not out of place among the normative sample.

Learning was not as strong in the Stroop task and there was little evidence for learning in backward digit span. The lowest levels of learning were seen in the semantic fluency task. The slope values for both total scores and first responses scores were not representative of the normative sample. Moreover, the absence of learning in this task was not readily explained in terms of fatigue in particular sessions. Figure 7.5 shows that performance was quite flat across all five sessions.

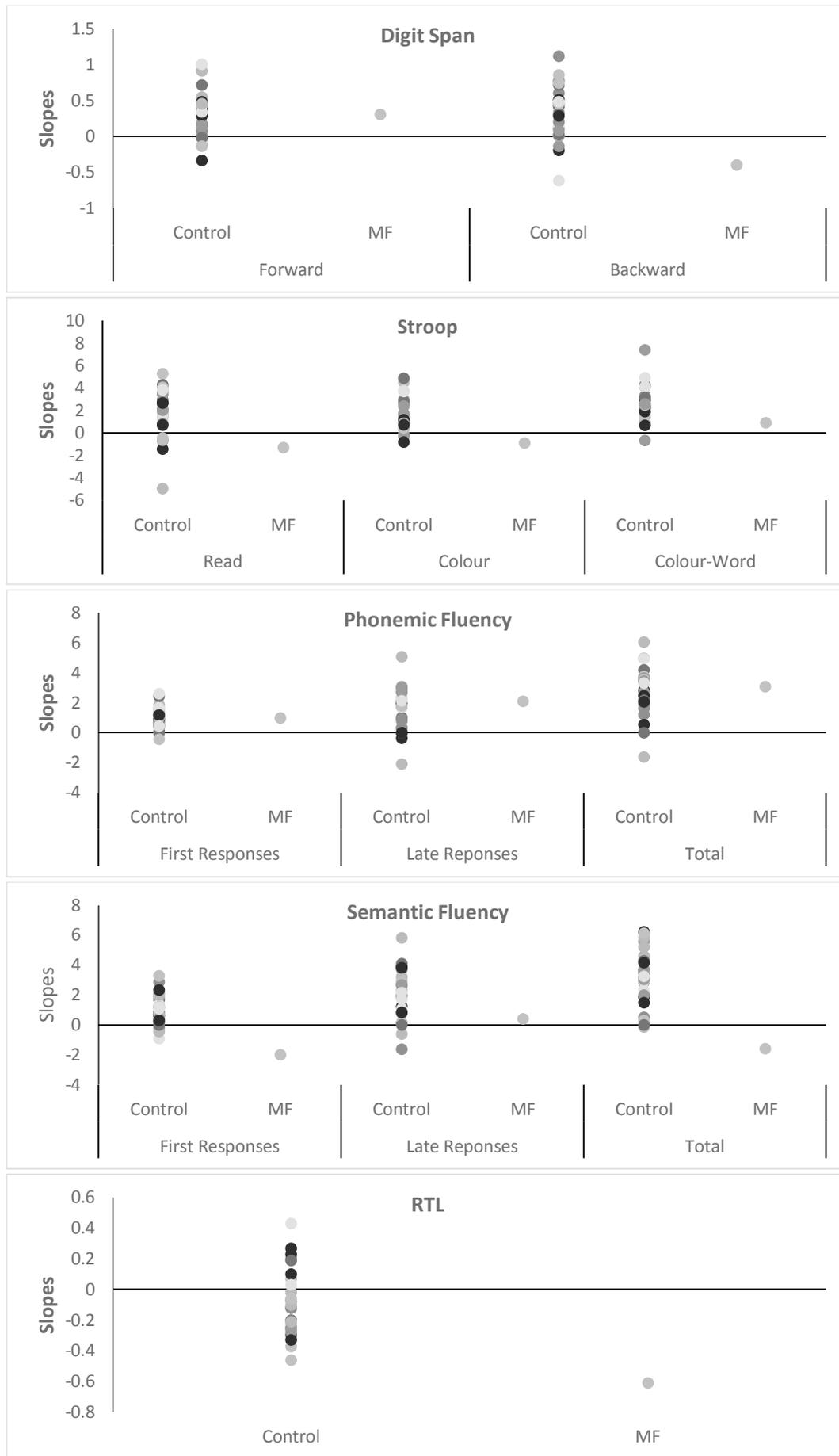


Figure 7.5. MF's slope values for all tasks compared to the normative sample.

Table 7.9

*MF's Learning Slopes Relative to the Normative Sample*

	Control	MF	Z	%ile
<b>Digit Span</b>				
Forward	.22 (-.42)	0.30	0.24	59
Backward	.28 (-.88)	-0.40	-1.14	12
<b>Stroop</b>				
Word	1.68 (-2.06)	-1.30	-1.36	9
Colour	1.59 (-1.41)	-0.90	-1.62	6
Colour Word	2.23 (-.35)	0.90	-1.01	16
<b>Phonemic Fluency</b>				
1st Responses	.97 (-.41)	1.00	0.04	51
Later Responses	1.74 (-1.19)	2.10	0.24	60
Total	2.71 (-.55)	3.10	0.23	60
<b>Semantic Fluency</b>				
1 <sup>st</sup> Responses	1.06 (-.65)	-2.00	-3.49	1
Later Responses	2.13 (-1.14)	0.40	-1.04	15
Total	3.31 (-.26)	-1.60	-2.70	1
Rey Tangled Lines	.22 (-.42)	0.30	0.24	60

Note: Numbers in parentheses in Control column represent the lower limit of 95% confidence interval.

### 7.8.3 Impaired and Unimpaired Tasks

MF's performance is summarised in Table 7.10 and her outcomes in the non-cognitive tasks can be characterised in terms of three groups. The first group represents those tasks where absolute levels of performance were within normal limits at baseline, and there

was positive learning across subsequent test sessions. Forward digit span, Rey Tangled Lines, and phonemic fluency tasks fell into this group. Performance on backward digit span and the first responses in the semantic fluency tasks were instances of the second response group, where initial levels of performance were within normal limits but there was little evidence of learning taking place. The third group contained Stroop and the later and total measures of the semantic fluency task. On these tasks, baseline performance was outside normative levels and there was again no evidence of learning on subsequent sessions.

Table 7.10

*Summary of Performance Outcomes for MF*

	Impairment	Learning	
		Conf. Intervals	Slopes
Digit Forward	N	?	Y
Digit Backwards	N	N	N
Phonemic - First	N	Y	Y
Phonemic - Later	N	Y	Y
Phonemic - Total	N	Y	Y
Semantic - First	N	N	N
RTL	N	Y	Y
Stroop - ColWord	Y	N	N
Stroop - Read	Y	N	N
Stroop - Colour	Y	N	N
Semantic - Later	Y	N	N
Semantic - Total	Y	N	N

MF's performance, collapsed across tasks is presented in Figure 7.6. Figure 7.6 suggests that on the unimpaired tests, MF showed standard practice effects across the first four sessions, with a down turn in the fifth session, where she indicated during testing that she was particularly fatigued. In the impaired tests, there was not consistent improvement even if one ignores the fifth session. Slopes were calculated on the basis of the first four sessions. MF's slope for the unimpaired tasks was .42 (SE = .95) and for the impaired tasks was .17 (SE = .13). The slope value for the unimpaired tasks was within the 95% confidence

intervals associated with normative performance, .48 [95% CI = .41, .57], but was outside on the impaired tasks.

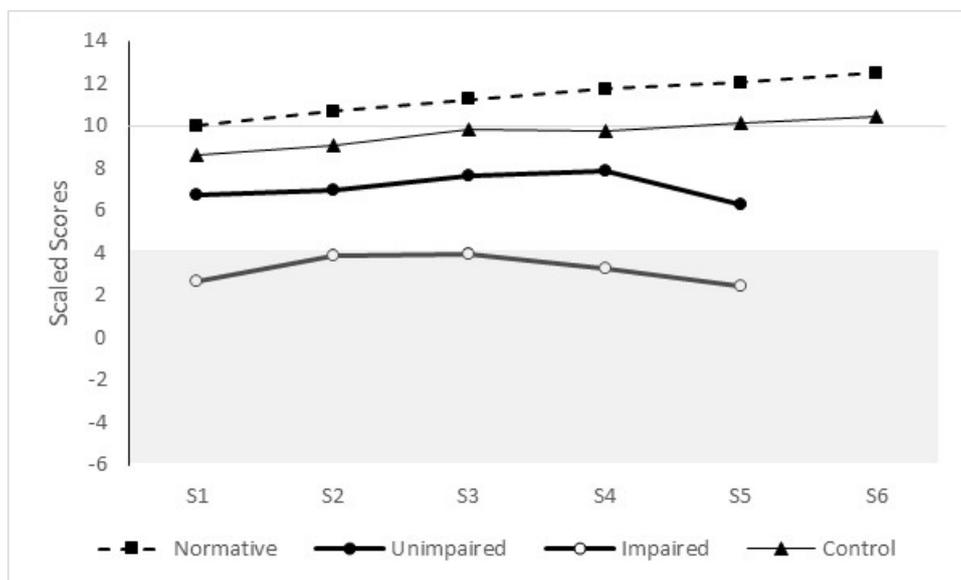


Figure 7.6. MF's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.

## 7.9 Discussion

MF was initially unimpaired on seven of the measures and showed normal learning on five of these tasks across repeated sessions. MF showed impairment on five of the primary measures, and there was limited cognitive recovery across these tasks. When collapsed across task, there were strong indicators of practice effects in the unimpaired domains across the first four sessions. The practice effects were of the same magnitude as those observed in the normative sample. This would indicate that MF shows normal practice effects in the unimpaired tasks.

The outcomes on the impaired tasks were not as positive. At both the individual task level and where performance has been averaged across tasks, there was consistent evidence that there had been no sustained improvement in performance, even when performance was limited to the first four sessions. In short while MF showed evidence of normal practice effects in the unimpaired tasks, she did not show evidence of recovery.

With regard to the diagnostic information that repeated testing may provide, MF's results confirm the necessity to distinguish between unimpaired and impaired tests.

Consideration of a subset of MF's scores would suggest that her cerebrovascular event has had little cognitive impact in that absolute scores were within normal limits, and standard practice effects were observed. The impact of the stroke is only evident on those tests where initial performance was outside normal range.

One final aspect of MF's performance is worthy of comment. Performance on Session 5 was consistently lower than on the other sessions, and she was aware when doing the testing that she was more fatigued than normal. These results confirm that obtaining stable data in the acute phase is problematic and suggests that as many testing sessions as possible should be conducted during this phase. Increased testing would highlight and hopefully reduce between-sessions variability that can occur through unwanted factors like fatigue.

## 7.10 Case Study 3: KS

### 7.10.1 Participant Background

KS is a 55-year-old male farmer from rural Queensland who was admitted to hospital following a Middle Cerebral Artery stroke in his right hemisphere.

**7.10.1.1 Precipitating event.** KS, a farmer from country Queensland, was holidaying overseas when he had sudden onset speech and motor deficits. Details surrounding actual events at the time are unclear, however reports indicate that KS was taken immediately to hospital where he received treatment before being transferred back to Australia. On admission to hospital in Brisbane, KS was densely hemiplegic on the left side of his body, causing left-side facial droop, and significant mobility impairment. When not in physiotherapy or occupational therapy, KS was wheelchair bound throughout his admission.

**7.10.1.2 Diagnosis.** KS was diagnosed with a right Middle Cerebral Artery Ischaemic stroke. Results from a speech and language assessment indicated KS was suffering from Dysphagia (difficulty swallowing) as a result of left-side facial droop (hemiplegia).

**7.10.1.3 Cognitive performance.** Premorbid estimates of KS's intellectual functioning indicated a FSIQ = 96 [90% CI 73, 119]. Little information was available regarding KS's current level of cognitive functioning, however observation suggested no obvious global or domain specific deficit 3 weeks after his stroke. KS demonstrated intact insight into his physical disabilities, but was anxious about recovery due to responsibilities back on his farm estate. Ward clinicians report that KS was extremely driven and motivated to make gains in therapy, and consistently worked hard during physiotherapy and occupational therapy.

## 7.11 Method

### 7.11.1 Test Administered

The tests were administered to KS five times. The tests that were administered on each session are summarised in Table 7.11. The tests were always administered in the same order as with the other patients, and were administered using standard instructions and standard procedure. Abbreviated instructions were used in the later test sessions.

Table 7.11

*Tests Administered to KS across five Test Occasions*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Digit Span</i>							
<i>Forward</i>	Single	✓	✓	✓	✓	✓	
<i>Backward</i>	Single	✓	✓	✓	✓	✓	
<i>Stroop</i>							
<i>Read</i>	Single	✓	✓	✓	✓	✓	
<i>Colour</i>	Single	✓	✓	✓	✓	✓	
<i>Colour-word</i>	Single	✓	✓	✓	✓	✓	
<i>Verbal Fluency</i>							
<i>Phonemic</i>	Single	✓	✓	✓	✓	✓	
<i>Semantic</i>	Single	✓	✓	✓	✓	✓	
<i>Rey Tangled Lines (RTL)</i>	Single	✓	✓	✓	✓	✓	

**7.11.2 Test Sessions**

The first test session was held 53 days after having suffered this stroke. He was administered the test battery on six occasions during the following 18 days until he was discharged from the rehabilitation ward. Table 7.12 presents information about the duration of each test session and the interval between successive testing sessions.

Table 7.12

*Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions*

	Testing Session					
Timing	S1	S2	S3	S4	S5	S6
Duration	55	60	50	55	45	45
Interval	-	3	2	4	4	3

### 7.11.3 Test Notes

As indicated in Table 7.1, KS's training regime departed from plan in two ways. Only five testing sessions were completed before KS was transferred to the long-term rehabilitation ward for continued treatment on his motor difficulties. Secondly, Snap was administered only once.

KS provided verbal and written consent to participate in this study. Testing sessions were held in KS's single hospital room regularly after lunch at 1pm. He was often absent during booked time slots, however when he was available he would participate. All testing sessions went smoothly and largely without problems.

There were two events that may have an impact on how KS's performance should be interpreted. As indicated in the participants section, KS was flown back to Australia with clear symptoms of paralysis on the left side of the body as well as problems with swallowing. KS was clearly aware of these difficulties and engaged wholeheartedly with physiotherapy. Before training began, KS sought clarity on how cognitive training would benefit him, as he did not want to waste time. The basics of cognitive rehabilitation were explained in the context of "physiotherapy for the brain", as KS very much enjoyed and engaged in physio. KS completed the first session of Snap but became agitated and couldn't understand how it would benefit him. The rest of the training session was completed smoothly, however KS reported that he felt it was beneath his level of cognitive functioning and he "felt like he was in preschool." This belief was gently confronted, with an explanation that the task instructions can seem easy but people often perform worse than expected, and may not be aware of any cognitive issues. KS agreed to continue testing but it was decided not to continue with the Snap task as the instructions to this task reinforced the view of doing "childish" tasks.

One further episode sheds some light on KS's attitudes to the testing regime. On at least three occasions during the phonemic fluency task he began one of the trials, and having exhausted his first responses he paused and then asked if enough words had been given to satisfy the assessor. He was reminded that the task was timed, and to continue until one minute had passed. One interpretation of this question is that there was minimal motivation to do the task well. The other is that such questions were asked to avoid having to acknowledge that he may have some cognitive issues.

KS appeared to have insight problems into the true nature of the effects of the stroke such that he demonstrated unrealistic expectations of recovery. KS said on a number of occasions that he expected to be physically able again by the end of his hospital stay to go back to working his farm. His severe disability caused by the stroke made this an unrealistic goal. According to hospital policy, KS would have been fully briefed on the severity of the stroke and the prognosis. KS appeared to discount such advice and the evidence of his impairment in the hope that he could return to his duties on the farm.

## **7.12 Results**

### **7.12.1 Absolute Performance**

**7.12.1.1 Initial levels of impairment.** KS's performance on the battery is summarised in Table 7.13 and Figure 7.7. KS presented as unimpaired on initial testing on 6 of the 14 primary measures. Specifically, KS was within normal range functioning on both digit span measures, all measures of phonemic fluency and on the Rey Tangled Lines test. However, on the Anagram task, the three components of the Stroop task, and the three measures of the semantic fluency task, KS's performance on the initial test was in the clinical range. Thus, KS presented with multiple deficits on first testing.

Table 7.13

*KS's Performance Across Tasks Expressed in Z-scores*

Tasks	Tests					
	1	2	3	4	5	6
<b>Digit Span</b>						
Forward	-0.28	0.16	0.30	0.37	0.30	
Backward	-0.81	-1.03	-1.30	-1.68	-1.86	
<b>Stroop</b>						
Read	<b>-5.12</b>	-4.13	-3.99	-3.08	-2.58	
Colour	<b>-2.91</b>	-2.65	-2.50	-1.88	-0.97	
Colour-Word	<b>-2.17</b>	-1.23	-1.53	-1.61	-1.38	
<b>Phonemic Fluency</b>						
1 <sup>st</sup> Responses	-0.01	-1.15	-1.67	-1.25	-1.08	
Later Responses	-1.49	-1.57	-2.05	-0.92	-1.21	
Total	-1.12	-1.60	-2.11	-1.16	-1.24	
<b>Semantic Fluency</b>						
1 <sup>st</sup> Responses	<b>-5.12</b>	-2.13	-1.89	-1.65	-2.21	
Later Responses	<b>-3.48</b>	-2.64	-4.09	-2.88	-3.04	
Total	<b>-3.83</b>	-2.52	-3.63	-2.70	-2.80	
Rey Tangled Lines	-1.88	0.57	0.25	0.62	-1.13	

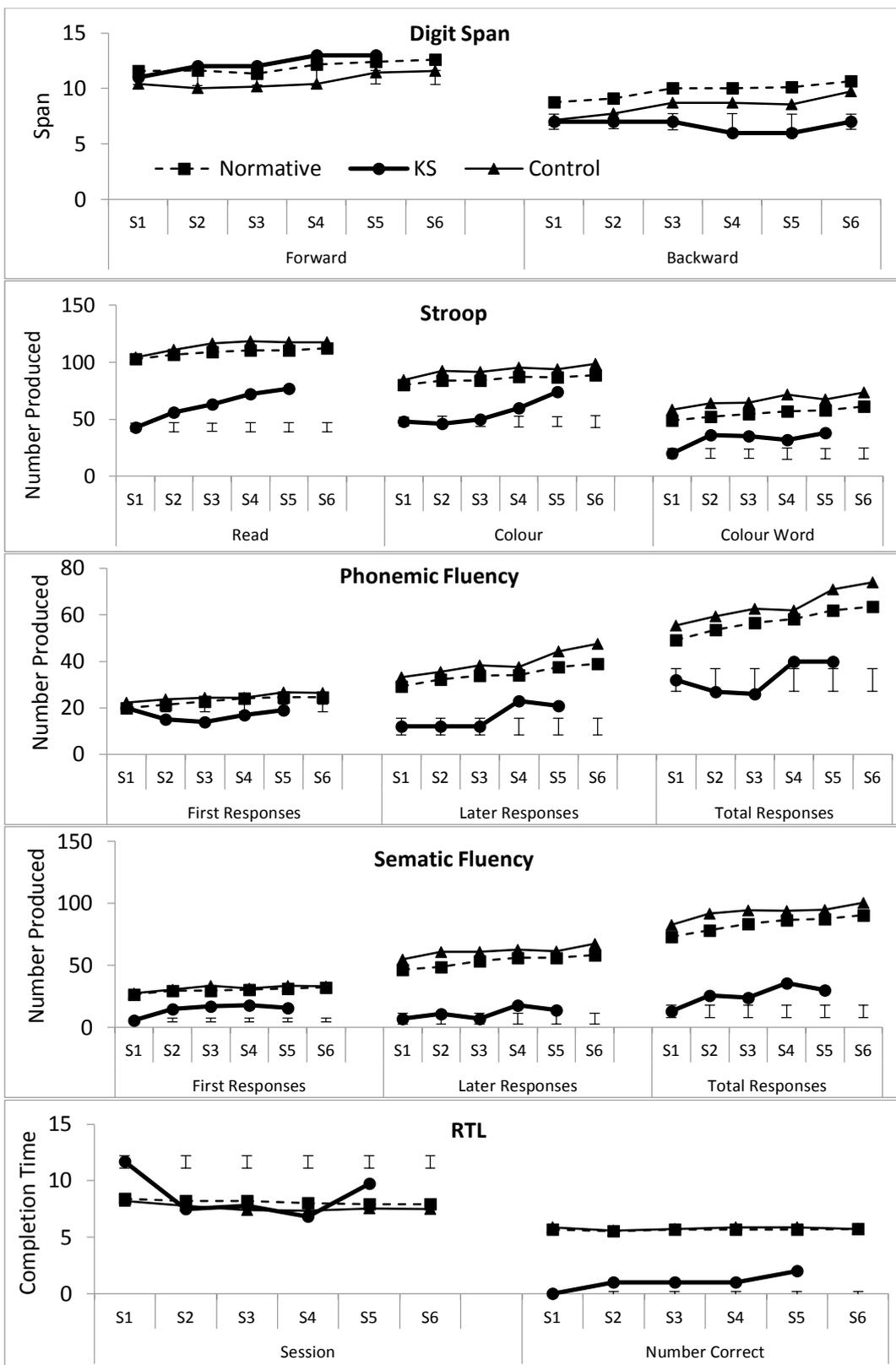


Figure 7.7. KS's performance relative to normative sample.

### 7.12.2 Learning

**7.12.2.1 Temporal stability confidence intervals.** As indicated earlier, half the measures were in the normal range on initial testing. In the case of forward digit span, the later responses and total responses on the phonemic fluency task significantly improved in a consistent way across the testing session, suggesting that task performance on these tasks was not affected by the stroke. In the case of backward span and the first responses in the phonemic fluency tests, performance remained within normal limits, but there was no evidence of systematic improvement in performance. In all sessions, performance varied within the 95% confidence intervals or was just outside those limits, but in the negative direction. In the case of the Rey Tangled Lines tests there were large fluctuations in performance across sessions, making it difficult to determine the presence of systematic changes. This was due to a single trial in each of Sessions 1 and 5 that took over 20 seconds to complete. With these two data points eliminated, KS's performance looks like the normative sample both in terms of completion time and in positive learning effects.

Performance on the Stroop, semantic fluency and Anagram tasks were all initially in the clinical range. All measures in the Stroop and semantic fluency tasks consistently improved across the sessions, such that by the fifth session scores on the colour naming, colour-word naming and semantic first responses were all within normal range functioning. In the case of the Anagram task there were once again large variations in performance that precluded the consistent improvement that was evident in the Stroop and semantic fluency tasks. However, by the fourth and fifth sessions, performance on this task had significantly improved above baseline measures.

In 8 of the 12 primary measures, there was evidence of consistent improvement across successive sessions. This is the pattern of performance that is expected if there is rapid recovery after stroke.

**7.12.2.2 Slopes.** KS's learning outcomes as measured by slope values are presented in Figure 7.8 and Table 7.14. The data complement that found using temporal stability confidence intervals, where KS showed positive learning effects across most measures on most tasks. His slopes were within normal bounds on all tasks and for most measures his learning was indistinguishable from the normative sample. The one exception to this general rule was that he exhibited little learning in backward digit span, and his first responses on the phonemic fluency task were also relatively low, but not in the clinical range. In short, KS demonstrated normal learning across all but two measures.

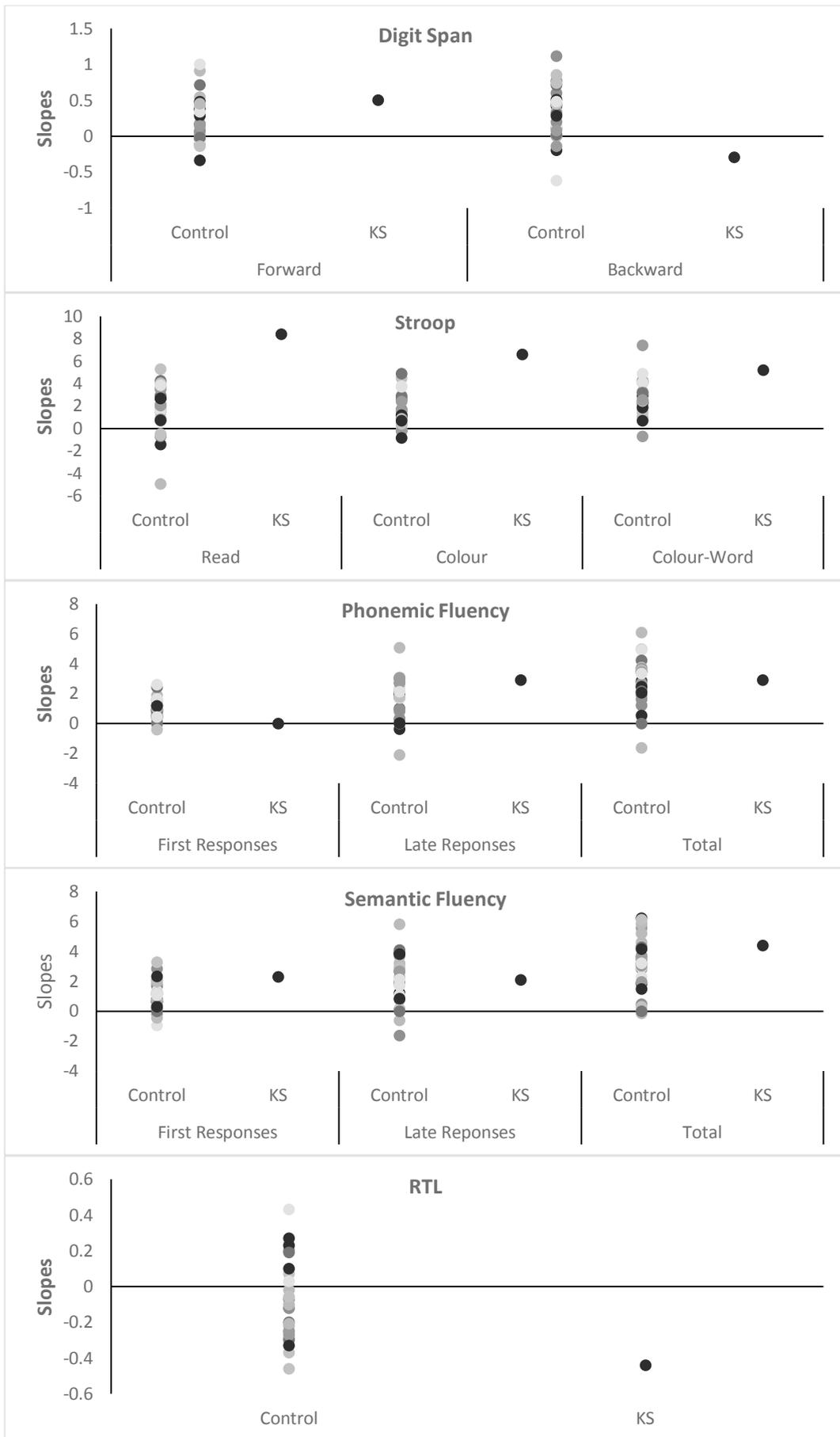


Figure 7.8. Slope values for KS relative to normative and control samples.

Table 7.14

*KS's Learning Slopes Relative to the Normative Sample*

Task	Control	KS	Z	%ile
<b>Digit Span</b>				
Forward	.22 (-.42)	0.50	0.86	80
Backward	.28 (-.88)	-0.30	-0.98	17
<b>Stroop</b>				
Word	1.68 (-2.06)	8.40	3.08	99
Colour	1.59 (-1.41)	6.60	3.27	99
Colour Word	2.23 (-.35)	5.60	2.54	99
<b>Phonologic Fluency</b>				
1st Responses	.97 (-.41)	0.00	-1.38	9
Later Responses	1.74 (-1.19)	2.90	0.78	78
Total	2.71 (-.55)	2.90	0.11	55
<b>Semantic Fluency</b>				
1 <sup>st</sup> Responses	1.06 (-.65)	2.30	1.41	93
Later Responses	2.13 (-1.14)	2.10	-0.02	49
Total	3.31 (-.26)	4.40	0.59	72
Rey Tangled Lines	-.08 (.35)	-0.44	-1.64	95

Note: Numbers in parentheses in Control column represent the lower limit of 95% confidence interval.

### 7.12.3 Impaired and Unimpaired Tasks

As Table 7.15 shows, KS's performance at baseline testing was impaired on half the measures used and he was within normal range functioning on the other half of the tests. Strong learning effects were present on most unimpaired measures and on those measures showing initial impairment. With the exception of backward digit span and the first responses

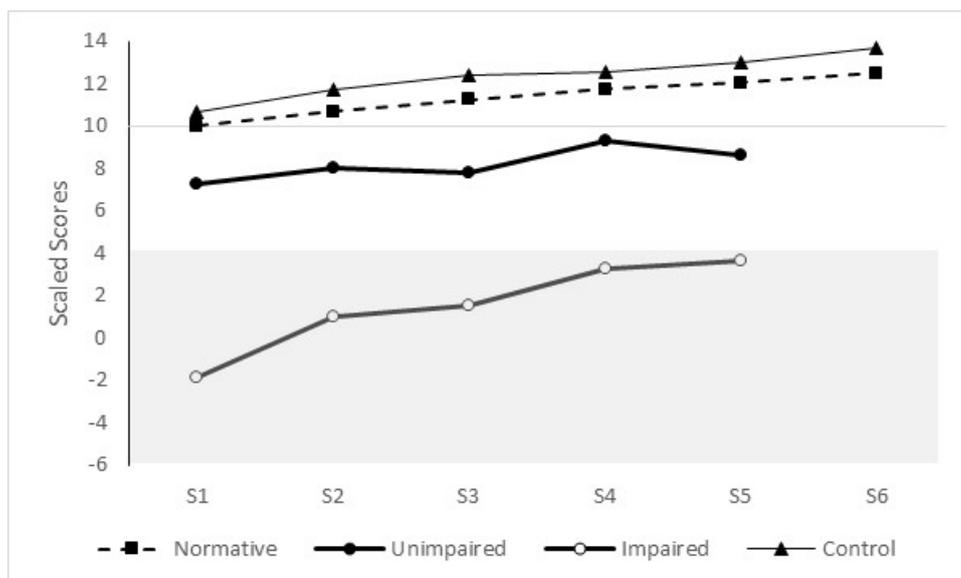
in the phonemic fluency task, there was sustained and consistent improvement across testing sessions. The other caveat involved the correct answers on the Rey Tangled Lines test, where frequent errors were made. It is not clear whether this represents a speed-accuracy trade-off or whether there are real problems in visuo-spatial abilities.

Table 7.15

*Summary of Performance Outcomes for KS*

	Impairment	Learning	
		Conf. Intervals	Slopes
Digit Forward	N	Y	Y
Digit Backwards	N	N	N
Phonemic - Later	N	Y	Y
Phonemic - Total	N	Y	Y
Phonemic - First RTLT	N	N	N
	N	Y	Y
Stroop - ColWord	Y	Y	Y
Stroop - Read	Y	Y	Y
Stroop - Colour	Y	Y	Y
Semantic - First	Y	Y	Y
Semantic - Later	Y	Y	Y
Semantic - Total	Y	Y	Y

KS's performance, collapsed across tasks, is presented in Figure 7.9. The Figure suggests that on the unimpaired tests, KS shows standard practice effects across the five sessions he completed. In the impaired tests, there was rapid improvement such that by the final test session performance was nearly in normative range. KS's slope for the unimpaired tasks was .40 (SE = .64) and for the impaired tasks was 1.32 (SE = .91). The slope value for the unimpaired tasks was just outside the 95% confidence intervals associated with normative performance, .48 [95% CI = .41, .57]. The value for the impaired task was well outside the upper limit of normative performance indicating greater levels of improvement on these tasks than exhibited by the normative sample.



*Figure 7.9.* KS's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.

### 7.13 Discussion

KS showed initial impairment on six of the primary measures. He did, however, show improved performance on all measures save digit span and the initial responses on the phonemic fluency task. However, improvements were not uniform across tasks. Figure 7.9 shows that on the initially unimpaired tasks improvements were similar in magnitude to those exhibited by the normative sample. KS produced near normal practice effects. In contrast, with the initially impaired tasks improvements were substantially greater than those observed in the normative sample. One plausible explanation for such rapid improvement in performance is that KS is showing recovery.

Again from a diagnostic perspective, KS's performance reinforces the need to distinguish between initially impaired and unimpaired tasks in determining the cognitive effects of stroke. KS's data also indicates that recovery over a short period can be observed in the weeks post-stroke.

## 7.14 Case Study 4: LL

### 7.14.1 Participant Background

LL is a 79-year-old male whose occupation was unknown, but who had less than 12 years of education. He was admitted to hospital with a stroke in his left hemisphere.

**7.14.1.1 Precipitating event.** LL was admitted to hospital for surgical removal of one of his kidneys and adrenal gland. On recovery, it was observed by hospital staff that LL had a right-sided facial droop and right sided peripheral weakness. His mobility was significantly impacted and he was wheelchair bound when not participating in therapy.

**7.14.1.2 Diagnosis.** LL suffered a stroke in the Corona radiata area of his left hemisphere producing, motor weakness on his right side, and swallowing and intelligibility issues.

**7.14.1.3 Cognitive performance.** Premorbid estimates of his intellectual functioning indicated a FSIQ = 94 [90% CI 73, 113]. On admission, LL's level of cognitive functioning was assessed using the MOCA, with scores indicating significant impairment (MOCA= 13/30). Specifically, LL demonstrated particular difficulty in visuo-spatial and attention domains, as well as a poor result on a verbal fluency task. Observations made by hospital staff indicate that LL had limited insight into his mobility limitations (high risk for falls), as he would continually forget to ask for assistance when leaving the bed.

**7.14.1.4 Other factors.** LL has a medical history including Parkinson's disease and Glaucoma.

## 7.15 Method

### 7.15.1 Test Administered

The tests were administered to LL six times. The tests that were administered on each session are summarised in Table 7.16. The tests were always administered in the same order as with other patients and were administered using standard instructions and standard procedure.

Table 7.16

*Tests Administered to LL across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<b>Digit Span</b>							
<i>Forward</i>	Single	✓	✓	✓	✓	✓	✓
<i>Backward</i>	Single	✓	✓	✓	✓	✓	✓
<b>Stroop</b>							
<i>Read</i>	Single	✓	✓	✓	✓	✓	✓
<i>Colour</i>	Single	✓	✓	✓	✓	✓	✓
<i>Colour-word</i>	Single	✓	✓	✓	✓	✓	✓
<b>Verbal Fluency</b>							
<i>Phonemic</i>	Single	✓	✓	✓	✓	✓	✓
<i>Semantic</i>	Single	✓	✓	✓	✓	✓	✓
<b>Rey Tangled Lines. (RTL)</b>	Single	✓	✓	✓	✓	✓	✓

### 7.15.2 Test Sessions

The first test session was held 22 days after having suffered this stroke. He was administered the test battery on six occasions during the following 16 days until he was discharged from the rehabilitation ward. Table 7.17 presents information about the duration of each test session and the interval between successive testing sessions.

Table 7.17

*Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions*

	Testing Session					
Timing	S1	S2	S3	S4	S5	S6
Duration	60	50	45	50	50	45
Interval		3	2	4	3	2

### 7.15.3 Test Notes

LL provided verbal and written consent to participate in this study. Training was completed over six sessions in LL's hospital room, which he shared with three other patients. He would complete tasks whilst sitting in his wheelchair, often after physiotherapy sessions in the morning. LL had an easy going personality and was happy to participate as long as it did not interfere with his physiotherapy appointments, which he very much enjoyed.

Instructions were explained to LL during the first session, and he was able to perform the verbal fluency, Stroop and Digit Span tasks without further prompting. He demonstrated difficulty with the "no finger tracing" instruction on the Rey Tangled Lines test, and required three reminders to use his eyes only. During the Anagram task, instructions were understood, however LL found it difficult to stay awake during this repetitive task and fell asleep four times. Similarly, the repetitive nature of the Snap task made attending and concentrating difficult, and LL demonstrated particular difficulty understanding the Snap instructions. Numerous methods of explanation and demonstration were used, however reliable data could not be gathered. The Snap task was not administered on subsequent sessions.

Instructions were explained to LL at the beginning of all subsequent sessions, however knowledge of the tests remained, and instructions were merely a reminder. There was often ambient noise in LL's room, causing distraction and interruption to training sessions.

## 7.16 Results

### 7.16.1 Absolute Performance

**7.16.1.1 Initial levels of impairment.** LL's absolute levels of performance across the six tests are presented in Figure 7.10 and Table 7.18. At the initial testing, LL's performance was outside the 95% confidence levels on 6 of the 12 primary measures. The aspects of impaired performance were in the Stroop read and colour naming, the three semantic fluency measures and the Rey Tangled Lines test. However, the fact that performance on digit span,

Stroop colour-word, and all three phonemic fluency measures were in the normative range may be somewhat misleading in that on the second session, both forward and backward recall and the later and total responses in the phonemic fluency tasks were in the clinical range.

### **7.16.2 Learning**

**7.16.2.1 Temporal stability confidence intervals.** Compared to initial test levels, LL showed little evidence of improvement across test sessions. In fact, digits forward, Stroop read, colour-word naming, phonemic fluency later responses, total scores and Rey Tangled Lines performance deteriorated across sessions, with deterioration particularly pronounced in the first three sessions. In the case of backward digit span, Stroop colour naming and first responses in phonemic fluency, there was no appreciable change across sessions. The only tasks where learning was evident was in the three semantic fluency tasks. However, even in this task the learning effects were not strong and performance came from a very low base.

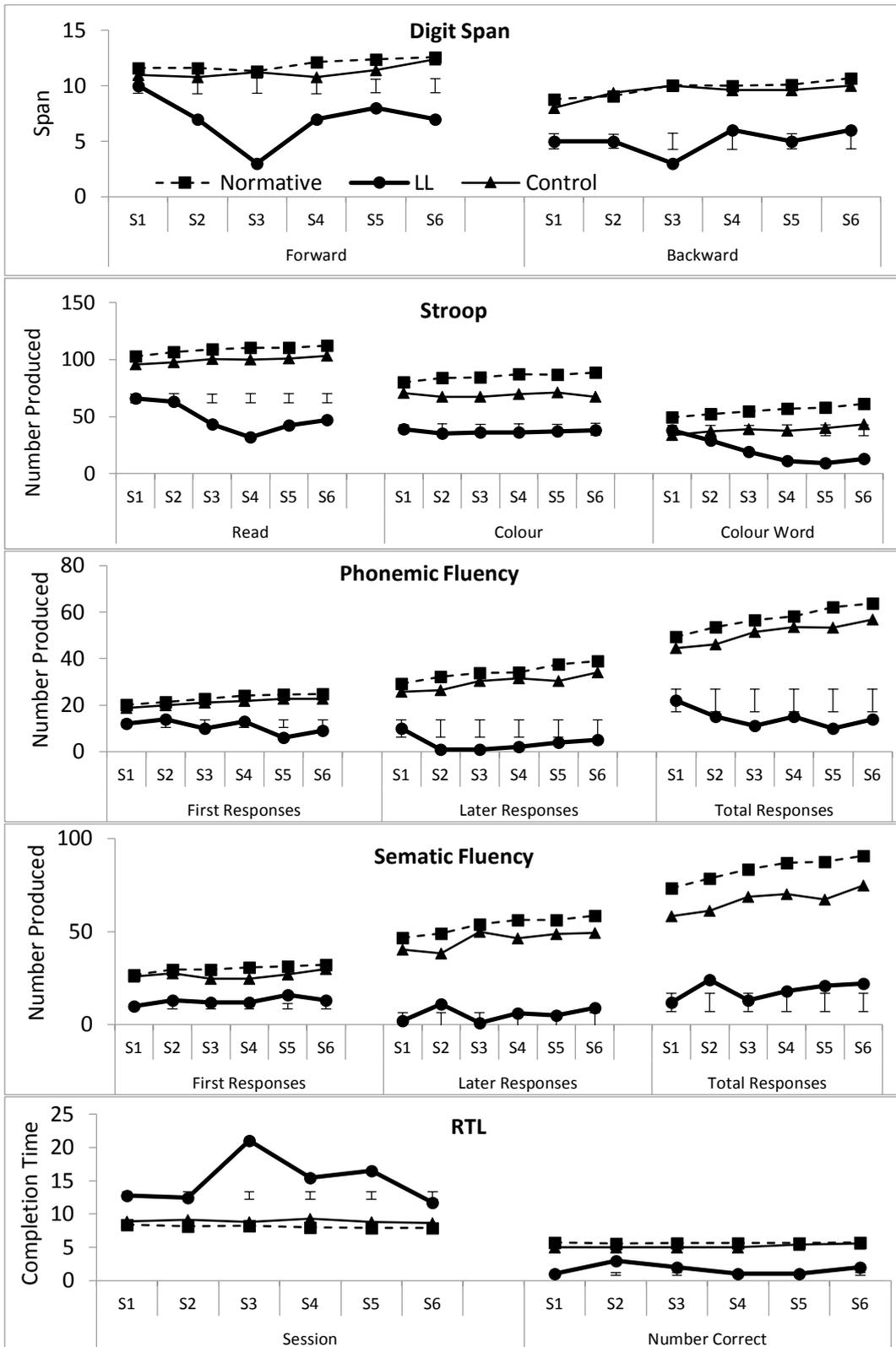


Figure 7.10. LL's performance relative to normative sample.

Table 7.18

*LL's Performance Across Tasks Expressed in Z-scores*

Task	Test Sessions					
	1	2	3	4	5	6
<b>Digit Span</b>						
Forward	-0.76	-2.02	-3.79	-2.22	-2.26	-2.84
Backward	-1.71	-2.02	-3.00	-1.68	-2.32	-2.19
<b>Stroop</b>						
Read	<b>-3.15</b>	-3.56	-5.73	-6.29	-5.30	-5.14
Colour	<b>-3.72</b>	-3.42	-3.52	-3.55	-3.77	-3.12
Colour Word	-0.83	-1.75	-2.79	-2.97	-3.39	-3.21
<b>Phonemic Fluency</b>						
1 <sup>st</sup> Responses	-1.58	-1.33	-2.43	-1.94	-3.56	-2.69
Later Responses	-1.66	-2.43	-3.09	-2.68	-2.46	-2.48
Total	-1.77	-2.32	-3.15	-2.75	-2.92	-2.81
<b>Semantic Fluency</b>						
1 <sup>st</sup> Responses	<b>-4.14</b>	-2.44	-2.67	-2.48	-2.21	-3.09
Later Responses	<b>-3.89</b>	-2.64	-4.60	-3.78	-3.67	-3.53
Total	<b>-3.89</b>	-2.62	-4.31	-3.67	-3.25	-3.54
Rey Tangled Lines	<b>-2.50</b>	-2.42	-7.70	-4.00	-5.40	-2.32

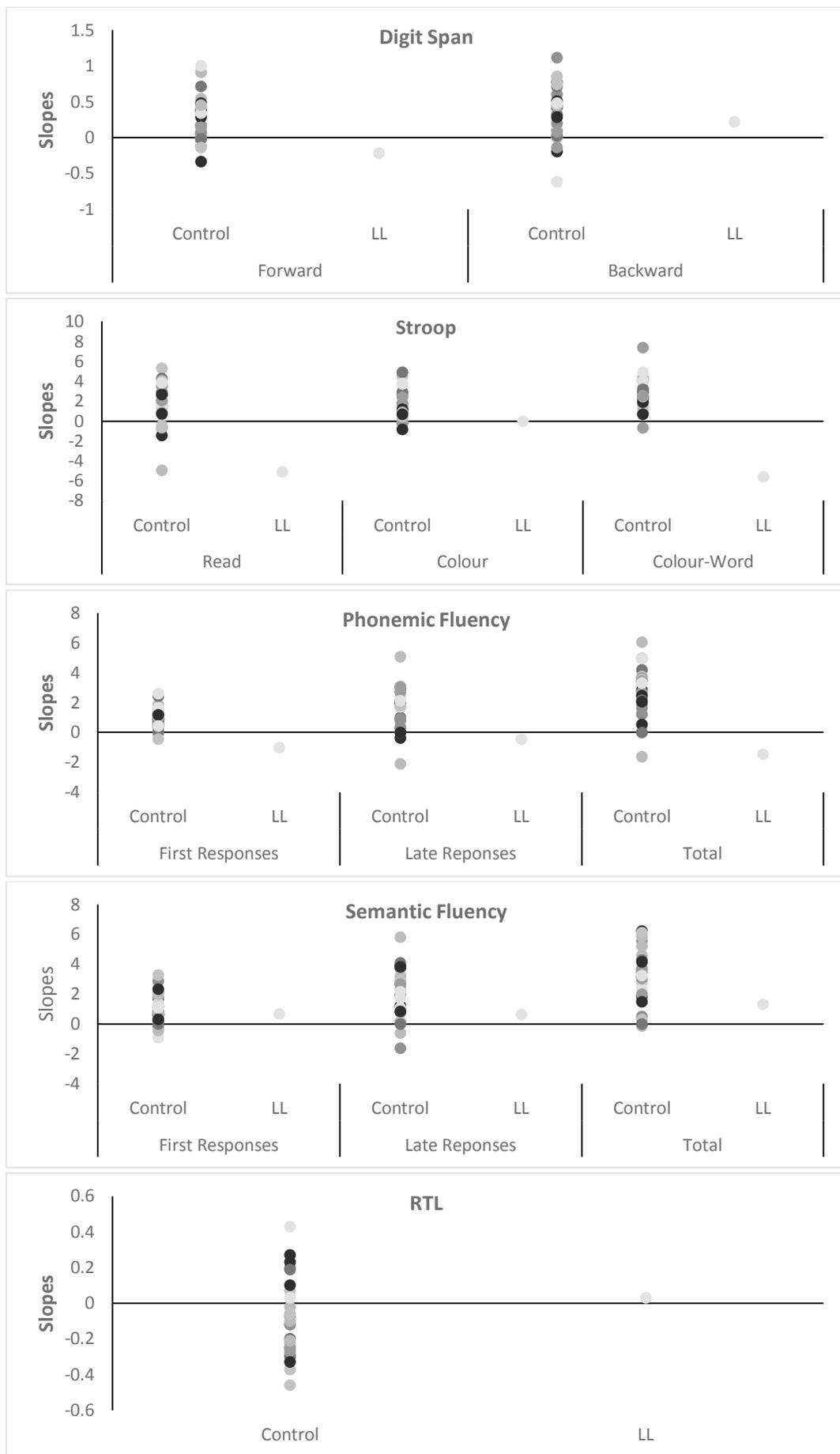


Figure 7.11. LL's slope values for all tasks compared to the normative sample.

Table 7.19

*LL's Learning Slopes Relative to the Normative Sample*

Task	Control	LL	Z	%ile
Digit Span				
Forward	.22 (-.42)	-0.22	-1.37	9
Backward	.28 (-.88)	0.22	-0.11	46
Stroop				
Word	1.68 (-2.06)	-5.11	-3.11	1
Colour	1.59 (-1.41)	0.00	-1.04	15
Colour Word	2.23 (-.35)	-5.63	-5.95	1
Phonologic Fluency				
1st Responses	.97 (-.41)	-1.02	-2.82	1
Later Responses	1.74 (-1.19)	-0.42	-1.44	8
Total	2.71 (-.55)	-1.45	-2.50	1
Semantic Fluency				
1 <sup>st</sup> Responses	1.06 (-.65)	0.68	-0.44	33
Later Responses	2.13 (-1.14)	0.62	-0.91	18
Total	3.31 (-.26)	1.31	-1.10	14
Rey Tangled Lines	-.08 (.35)	0.03	-0.49	32

Note: Numbers in parentheses in Control column represent the lower limit of 95% confidence interval.

**7.16.2.2 Slopes.** Learning effects are summarised in Table 7.19 and Figure 7.11. This measure of learning presented a similar pattern to that obtained using temporal stability confidence intervals. That is, there was very little evidence of any learning in any of the tasks. Many of the measures again showed negative slope values reflecting deterioration in performance rather than improvement. Consequently, it was not surprising that percentile

rankings were very low on these tasks. The only tasks that did show positive slopes were backward digit span and the three semantic fluency measures. However, even in these tasks, the learning effects were not overly strong.

### 7.16.3 Impaired and Unimpaired Tasks

Table 7.20

*Summary of Performance Outcomes for LL*

Tasks	Impairment	Learning	
		Conf. Intervals	Slopes
Anagram	N	N	-
Digit Forward	N	N	N
Digit Backwards	N	N	N
Stroop - ColWord	N	N	N
Phonemic - Later	N	N	N
Phonemic - Total	N	N	N
Phonemic - First	N	N	N
Stroop - Read	Y	N	N
Stroop - Colour	Y	N	N
Semantic - First	Y	N	N
Semantic - Later	Y	?	N
Semantic - Total	Y	Y	Y
RTL	Y	N	N

The effects of LL's stroke were quite widespread in terms of the range of tasks that were adversely affected, as is evident in Table 7.20. LL's strongest area of performance was on the Digit Span task, but even on this task performance was outside normal limits in many instances. Likewise, the semantic fluency task scores were outside the normal range, but LL did show some improvement across the later testing sessions. The level of improvement, however, was relatively weak. LL did not perform at all well on phonemic fluency and Rey Tangled Lines tests and performance deteriorated across sessions. There was no indication of any recovery in these tasks.

LL's performance, collapsed across tasks, is presented in Figure 7.12 and confirms the basic outcomes present in the individual tests. LL's slope for the unimpaired tasks was

-.37 (SE = .30) and for the impaired tasks was -.05 (SE = .20). The slope values for both sets of tasks were well outside the 95% confidence intervals associated with normative performance, .48 [95% CI = .41, .57]. The negative value on each slope indicates that performance deteriorated across the sessions.

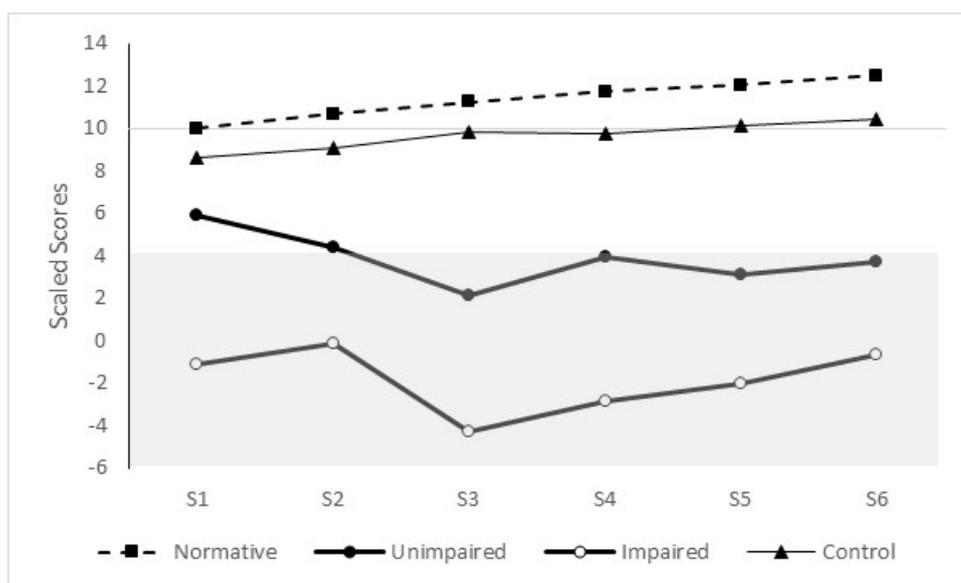


Figure 7.12. LL's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.

### 7.17 Discussion

LL showed impairment on six of the primary measures. However, the need to distinguish between impaired and unimpaired tasks was not essential as there was little evidence of learning in any of the measures. Thus, there are no practice effects and no recovery. However, there are some regularities in the outcomes that were also present in GL's data. Performance changes were of a similar magnitude for unimpaired and impaired tasks as reflected in near parallel curves. In the case of GL the two lines were relatively flat across all sessions, but with LL there appeared to be deterioration over the first three sessions, but slow improvement on the remaining three sessions.

## 7.18 Case Study 5: TB

### 7.18.1 Participant Background

**TB** is an 80-year-old, male, retired engineer who was admitted to hospital with a diagnosis of a left side cortical stroke.

**7.18.1.1 Precipitating event.** TB was unable to recall events of his stroke, however collateral information from his wife and the ambulance service indicate that TB collapsed but did not lose consciousness. He presented with slurring of speech and left sided weakness. TB reported feeling weak and having episodes of vomiting.

**7.18.1.2 Diagnosis.** Medical scans revealed that TB suffered damage to the left medial temporal lobe, left medial occipital lobe and left posterior thalamic areas. As a result of his stroke, TB had difficulty identifying objects in his right visual field.

**7.18.1.3 Cognitive performance.** Premorbid estimates of his intellectual functioning indicated a FSIQ = 113 [90% CI 90, 136]. On admission TB scored 26/30 (low range normal) on the MMSE. A day later he was re-tested and scored 17/30 (moderate cognitive impairment). Further investigation using the MOCA (11/30) identified deficits in the areas of language production, visuospatial tasks, recall and orientation. Speech therapy assessment identified evidence of mild receptive aphasia and mild to moderate expressive aphasia.

**7.18.1.4 Other factors.** TB had a history of depression, early cognitive impairment and had been diagnosed with Parkinson's disease in its early form.

## 7.19 Method

### 7.19.1 Test Administered

The battery was administered to TB on five occasions. The tests that were administered on each session are summarised in Table 7.21. Because of vision problems only the following tests were administered, and they administered in the same order: Phonemic fluency, semantic fluency, Digit Span, and the Stroop test. During the initial test session, standardised instructions were read and the tasks were administered in the standard way.

Table 7.21

*Tests Administered to TB across Five Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Digit Span</i>							
<i>Forward</i>	Single	✓	✓	✓	✓	✓	
<i>Backward</i>	Single	✓	✓	✓	✓	✓	
<i>Stroop</i>							
<i>Read</i>	Single	✓	✓	✓	✓	✓	
<i>Colour</i>	Single	✓	✓	✓	✓	✓	
<i>Colour-word</i>	Single	✓	✓	✓	✓	✓	
<i>Verbal Fluency</i>							
<i>Phonemic</i>	Single	✓	✓	✓	✓	✓	
<i>Semantic</i>	Single	✓	✓	✓	✓	✓	
<i>Rey Tangled Lines</i>	Single	✓					

### 7.19.2 Test Sessions

The first test session was held 22 days after having suffered his stroke. He was administered the test battery on five occasions during the following 28 days until he was discharged from the rehabilitation ward. Table 7.22 presents information about the duration of each test session and the interval between successive testing sessions.

Table 7.22

*Duration of Testing Sessions (Mins) and Intervals between Tests (days) for Training Sessions*

Timing	Testing Session					
	S1	S2	S3	S4	S5	S6
Duration	30	20	30	20	30	-
Interval		2	19	3	2	-

### 7.17.20 Test Notes

*Session 1:* TB provided written and verbal consent to participate in the study. TB was able to complete verbal and semantic fluency tasks but he was aware he was performing

poorly and was visibly frustrated. TB was reassured that practice might help and the training was designed for him to recover his abilities. Digit Span forwards and backwards were completed and TB performed well. The Stroop tasks were administered and while TB initially had difficulty reading the words, he moved the page on an angle so it was comfortable for him to read and was able to complete the tasks. TB attempted the Anagram task however he was unable to decipher the words. For example “STAITON” (*station*) was presented on an angle to see if he could see properly, but he reported he could not and guessed the word “Saturday”. This significant difficulty is likely due to a combination of his diagnosed visual disorder and language production difficulties. Similarly, Rey Tangled Lines was attempted, but the complex visual stimuli were too difficult to see, especially with the page manipulated on an angle. TB also tried the Snap task but by the end of the session he was fatigued and found it extremely difficult. As a consequence of these difficulties, only the digit span, Stroop, phonemic and semantic fluency tasks were used in subsequent sessions.

*Session 2:* TB presented as tired but was happy to start the session. Phonemic fluency and the naming Animals and Countries trials of the semantic fluency task were completed before TB became too fatigued to continue. The session recommenced the following day when TB was feeling energetic. He performed well on the remaining category of Fruits and Vegetables, digit span and the Stroop tasks.

TB became ill for a period of 2 weeks between Sessions 2 and 3. TB’s symptoms of dizziness and fatigue were investigated and he was diagnosed with Bradycardia. During this time TB was unable to participate in the testing program.

*Session 3:* On recommencing testing TB did not remember previous testing sessions but was happy to participate. He showed frustration with his difficulty in producing words during the two verbal fluency tasks. TB completed digit span forwards and backwards confidently and made minimal errors during the Stroop tasks.

*Session 4:* TB again showed frustration with word finding on the verbal fluency task and would often begin a word but could not finish it (e.g. Anti...). TB persevered until he was able to finish the word “antidote.” Again, the Digit Span and Stroop tasks were completed without frustration.

*Session 5:* The same pattern occurred with frustration expressed on the fluency tasks but ready completion of the Digit Span and Stroop tasks.

## 7.20 Results

### 7.20.1 Absolute Performance

**7.20.1.1 Initial impairments.** Given that TB’s problems with visual perception limited the number of tasks he could do, the results are limited to those tasks for which TB was able to complete multiple sessions. Thus, TB’s absolute levels of performance on the Digit Span, Stroop and verbal fluency tasks are presented in Table 7.23 and Figure 7.13. At initial testing, TB’s performance was outside the 95% confidence levels on 9 of the 11 primary measures. The specific impairments were in the read, colour naming and colour-word naming components of the Stroop task, and all measures associated with the phonemic and semantic fluency tasks. There were no initial impairments in Digit Span task.

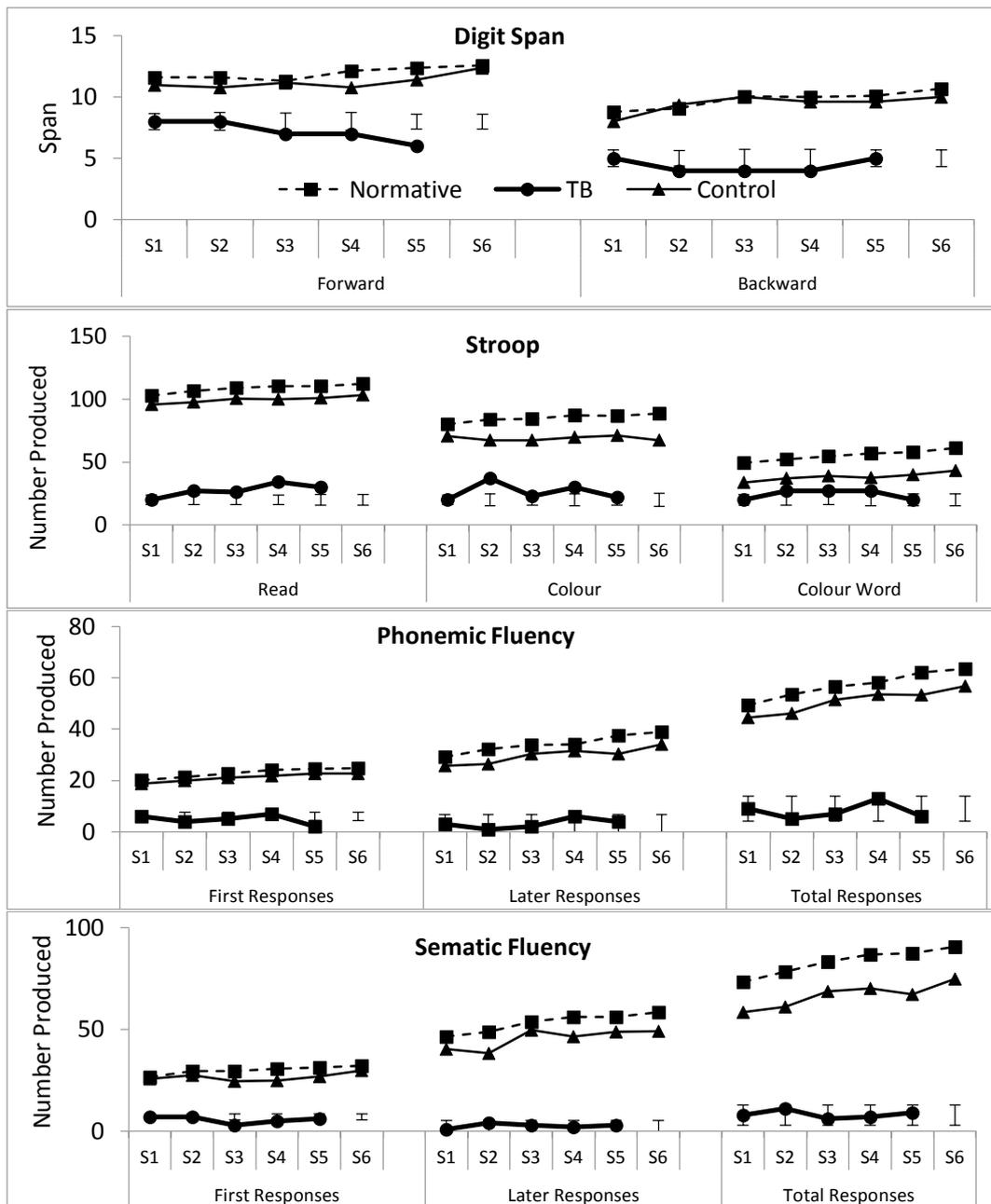


Figure 7.13. TB's performance relative to normative sample and control group.

Table 7.23

*TB's Performance Across Tasks Expressed in Z-scores*

Task	Test Sessions					
	1	2	3	4	5	6
<b>Digit Span</b>						
Forward	-1.71	-1.59	-1.97	-2.22	-3.29	
Backward	-1.71	-2.51	-2.57	-2.51	-2.32	
<b>Stroop</b>						
Read	<b>-7.09</b>	-6.50	-7.22	-6.13	-6.23	
Colour	<b>-5.44</b>	-3.28	-4.47	-3.96	-4.90	
Colour Word	<b>-3.65</b>	-1.90	-2.16	-1.93	-2.63	
<b>Phonemic Fluency</b>						
1 <sup>st</sup> Responses	<b>-2.75</b>	-3.13	-3.39	-2.98	-4.33	
Later Responses	<b>-2.27</b>	-2.43	-3.00	-2.34	-2.46	
Total	<b>-2.61</b>	-2.92	-3.43	-2.87	-3.15	
<b>Semantic Fluency</b>						
1 <sup>st</sup> Responses	<b>-4.88</b>	-3.39	-4.08	-3.45	-3.71	
Later Responses	<b>-3.97</b>	-3.11	-4.43	-4.08	-3.80	
Total	<b>-4.15</b>	-3.26	-4.74	-4.27	-3.84	

## 7.20.2 Learning

**7.20.2.1 Temporal stability confidence intervals.** Inspection of Figure 7.13 indicates that there was no consistent improvement on either of the Digit Span tasks, nor on any of the measures on the phonemic and semantic fluency tasks. In all instances, performance

fluctuations across test sessions fell within the confidence intervals, or just outside those intervals but in the negative direction.

In the case of the Stroop task, there was an initial improvement on the second test session for all measures and this improvement was maintained through Sessions 3 and 4. However, performance returned to within the test confidence band on the colour naming and colour-word naming on Session 5. In contrast, performance on the read component increased consistently across sessions.

**7.20.2.2 Slopes.** TB's slope measures are depicted in Figure 7.14 and Table 7.24. There was very little evidence of learning going on in any of the tasks, with the exception of the reading sub-test of the Stroop task. In all other tasks, learning was positive but weak or there was no learning occurring. The percentile values associated with each task confirmed that very little learning occurred on most measures.

One possible reason for the absence of such learning is that this was an artefact of the 2-week hiatus in testing when TB was ill. If this was the cause, one might expect to see an improvement from Session 1 to Session 2, a regression in Session 3 followed by improvement in Sessions 4 and 5. There were no strong patterns of this type in Figure 7.13. Certainly there was no strong evidence of increases in learning over Sessions 3 to 5.

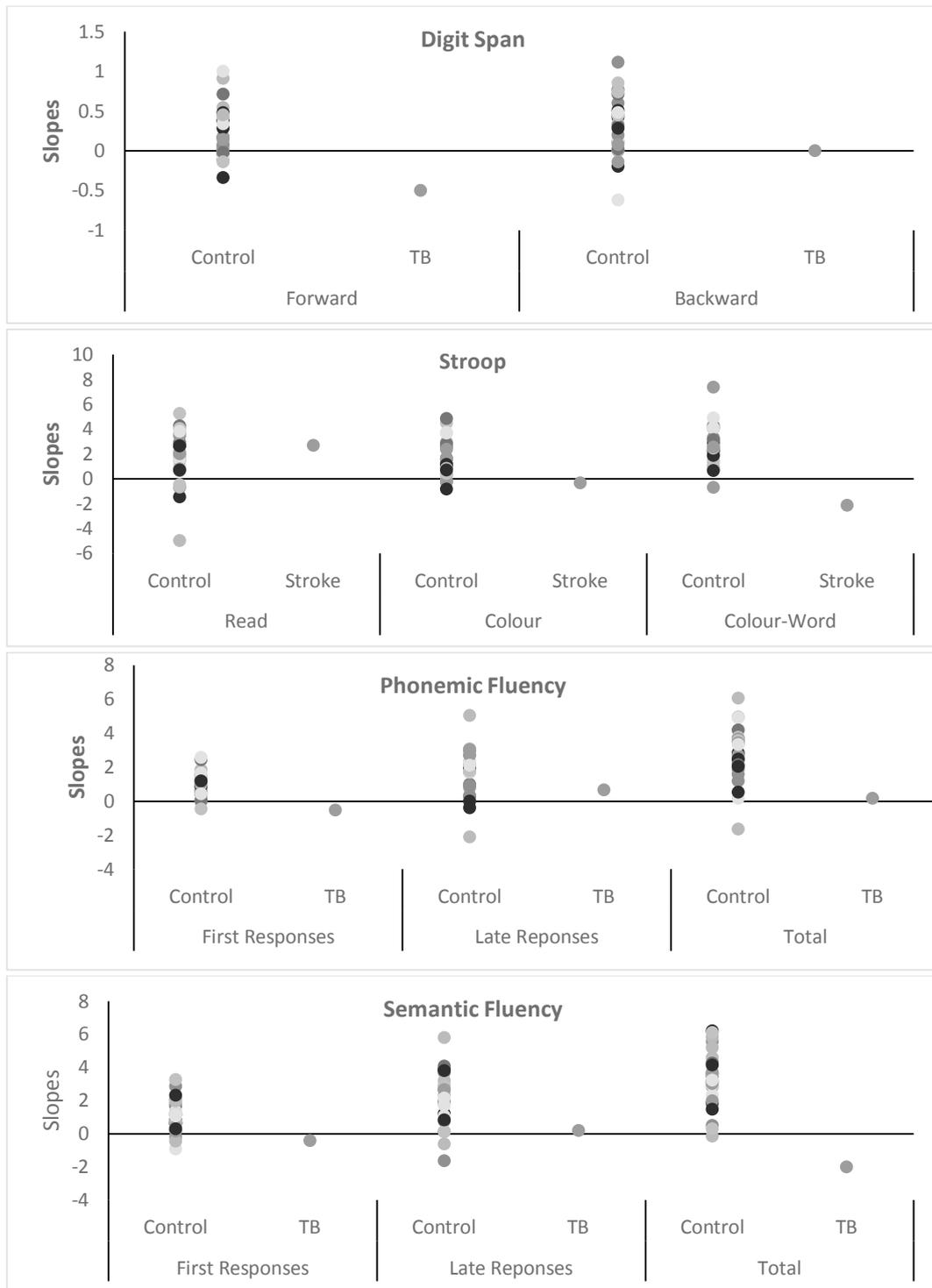


Figure 7.14. TB's slope values for all tasks compared to the normative sample.

Table 7.24

*TB's Learning Slopes Relative to the Normative Sample*

	Control	TB	Z	%ile
<b>Digit Span</b>				
Forward	.22 (-.42)	-0.50	-2.23	1
Backward	.28 (-.88)	0.00	-0.47	32
<b>Stroop</b>				
Word	1.68 (-2.06)	2.70	0.47	69
Colour	1.59 (-1.41)	-0.30	-1.23	11
Colour Word	2.23 (-.35)	-2.10	-3.28	1
<b>Phonemic Fluency</b>				
1st Responses	.97 (-.41)	-0.50	-2.08	2
Later Responses	1.74 (-1.19)	0.70	-0.69	25
Total	2.71 (-.55)	0.20	-1.51	7
<b>Semantic Fluency</b>				
1 <sup>st</sup> Responses	1.06 (-.65)	-0.40	-1.67	5
Later Responses	2.13 (-1.14)	0.20	-1.16	13
Total	3.31 (-.26)	-2.00	-2.92	1

Note: Numbers in parentheses in Control column represent the lower limit of 95% confidence interval.

### 7.20.3 Impaired and Unimpaired Tests

Table 7.25 summarises TB's performance on the tests he was able to complete. In terms of absolute levels of performance he was outside the 95% confidence interval on all tasks for most sessions. This would suggest that TB was impaired in all cognitive domains

covered by the four tests that he could complete. Having said this, there were indicators of normal cognitive processing in that forward digit span was better than backward span, and that in both fluency tasks initial responses were more forthcoming than later responses. However, unlike the other groups, he processed the same number of words across the three components of the Stroop task and production of instances in the semantic fluency task was no greater than that in the Phonemic fluency task.

Table 7.25

*Summary of Performance Outcomes for TB*

	Impairment	Learning	
		Conf. Intervals	Slopes
Digit Forward	N	N	N
Digit Backwards	N	N	N
Stroop - ColWord	Y	N	N
Phonemic - Later	Y	N	N
Phonemic - Total	Y	N	N
Stroop - Read	Y	Y	Y
Stroop - Colour	Y	N	N
Phonemic - First	Y	N	N
Semantic - First	Y	N	N
Semantic - Later	Y	N	N
Semantic - Total	Y	N	N

TB's impairment extended to learning in that there was very little evidence of improved performance across any of the measures used. The only task on which there was learning was on the reading component of the Stroop task. Of all the tasks used in the current battery, this is the one task that is overlearned and automatic. The general lack of improvement is reflected in Figure 7.15 where TB's performance has been collapsed across tasks. TB's slope for the unimpaired tasks was  $-.36$  ( $SE = .80$ ) and for the impaired tasks was  $-.01$  ( $SE = .01$ ). The slope values for both sets of tasks were well outside the 95% confidence

intervals associated with normative performance, .48 [95% CI = .41, .57]. The negative value on each slope indicates that performance deteriorated across the sessions.

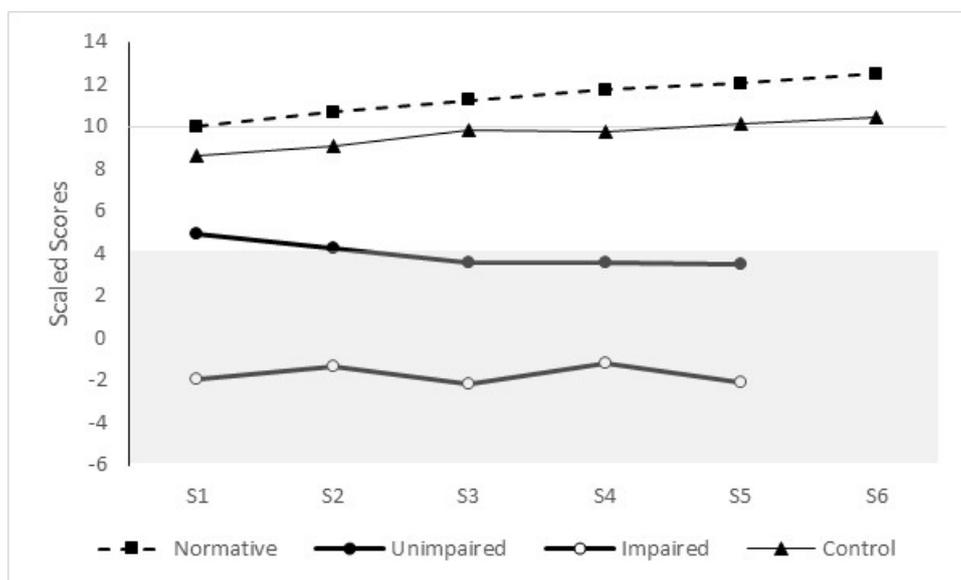


Figure 7.15. TB's performance on initially impaired and initially unimpaired tasks in comparison to normative and control groups.

### 7.21 Discussion

The visual problems that resulted from his stroke meant that the Anagram and Rey Tangled Lines test could not be administered. In spite of this, it was still possible to assess TB's performance on the remaining tests. In most respects the outcomes were quite similar to those produced by GL and LL. There was little evidence for learning on either impaired or unimpaired tasks such that there were no practice effects and no recovery. Like the other two patients with stroke, changes across sessions were of a similar magnitude for unimpaired and impaired tasks. Both decline in a similar manner.

### 7.22 General Discussion

The current case studies examined changes in performance of five patients with stroke in the neuropsychological tests in the test battery. Given that testing was initiated in the acute phase of stroke, the expectation was that on first testing, it was likely that patients would

show significant impairments on a range of tests, but that some domains would be unaffected by the stroke. That is, the patients with stroke would show multiple deficits rather than single deficits. However, the major interest centred on how performance would change over repeated assessment over a 2- or 3-week period while the patient was still in hospital. The expectation was that performance would parallel that of the normative sample in the domains that were unaffected by the stroke. The key issue was the extent to which performance would improve on the tests that showed impairment at initial testing.

With regards to the number of tasks in which there was initial impairment, expectations were met in that all five patients showed impairments on multiple tasks and measures. Of the 12 primary measures available, four of the patients showed a deficit on roughly half the measures with TB showing impairment on 9 out of 11 measures. Thus, the current data confirmed that the effects of a stroke were widespread in the early days and weeks following the stroke. The results also confirmed that multiple cognitive domains were affected but the specific domains, tasks or measures that were affected changed from person to person. Having said that, the Stroop task and the semantic fluency task were problematic for all patients whereas there was no initial impairment on either of the Digit Span tasks for any of the participants.

Changes in performance were examined both at the individual task level and when composite scores were generated for initially impaired and initially unimpaired tests. The use of temporal stability and slope estimates were two within-subject measures designed to assess the extent to which performance improved with repeated exposure at the individual test level. Both measures tended to complement each other, providing a relatively consistent index of learning. Converting the raw data to scaled scores allowed for across task composites to be created in order to make more global comparisons between those tasks that were initially impaired and those that were not. Using these measures, three distinct profiles emerged. KS

proved to be an example of the best-case outcome, to the extent that he showed improved performance on almost all measures on every task, those that were initially impaired and those that were initially unimpaired. The second profile is represented by MF who showed improvement in the tasks that were unimpaired at first testing, but showed no consistent improvement on the impaired tests. The third profile is that exhibited by GL, LL and TB. There was no consistent improvement on either unimpaired or impaired tasks. In short, the profiles range from a best-case scenario where there is improvement across all tasks to a near worst-case scenario where there is no evidence for learning on any of the tasks or measures.

The composite measures also allowed more fine-grained distinctions to be made. In the case of GL and KS, performance on the unimpaired tasks improved across training sessions, and their improvement was of a similar magnitude to that displayed by the normative (and control) group. Thus, normal practice effects were observed in the case of these two patients. Natural recovery, as opposed to or in addition to practice effects, was observed in KS's performance on the initially impaired tasks. KS's rate of improvement across sessions was much greater than that observed in the normative sample. The third facet of the data appeared to reflect a generalised learning deficit. GL, LL and TB all failed to show improvement on either initially impaired or initially unimpaired tasks. At the composite level, the curves/lines were roughly parallel for both impaired and unimpaired tasks despite across session fluctuations in performance. This result suggests that the learning slope can provide an additional source of evidence regarding the effects of stroke on the learning process that is independent of absolute performance on the task. As Darby et al., (2002; Duff et al., 2012) have suggested these learning effects can provide diagnostic information that is rarely used in the evaluation of cognitive performance.

The lack of improvement across sessions on all tasks in three of the patients with stroke was interpreted in terms of a generalised learning deficit. However, an alternative

explanation for the lack of learning is that between-session variability is such that it swamps any learning effects present. As reported in Chapter 2, many studies exploring the effects of stroke on neuropsychological functioning explicitly do not test people during the acute phase of stroke precisely because the effects of illness, fatigue, the need for sleep, and the ability to maintain concentration are all assumed to prevent an accurate and reliable assessment of performance (Jaillard et al., 2009; Nys et al., 2005, 2007). The effects of these variables do have an effect on performance in the current case studies. For example, MF reported that she was particularly fatigued in her fifth session and the effects can be seen in most of the individual tasks, and in the composite measures as reflected in a decrease in performance from Session 4. In addition, some of the tasks were more prone to momentary changes that resulted in large discrepancies between sessions. Digit span and the Rey Tangled Lines test were two clear examples of this (GL and LL). The benefit of repeated presentation of such tests is that those outlying data points can be ignored to some extent. Thus, while there were clear fluctuations in performance across sessions, it is possible to abstract a general trend in performance that reflects the true changes in performance across sessions, or more accurately, the lack of improvement across sessions.

In short, the repeated administration of a battery of neuropsychological tests across a 2- to 3-week period produced diverse outcomes among the patients involved in terms of the number and distribution of cognitive impairments. MF and KL showed normal practice effects on tasks that were originally within the normal range of functioning. However, any signs of natural recovery in impaired function during the acute phase were limited to KS. There was little evidence for recovery in the remaining four patients with three of them showing a generalised learning deficit.

## Chapter 8 Cognitive Tasks.

### 8.1 Introduction

The previous chapter described performance on the neuropsychological tasks. The current chapter centres on performance on the cognitive tasks. The original conceptualisation of the cognitive component was designed to be complementary to the neuropsychological tests in that there were four of each type. However, two of the tasks, the Common Associates Test and the Immediate Serial Recall tasks failed to meet the criteria for continued inclusion in the test battery and were therefore not administered to the patients with stroke. Consequently, the cognitive tests in the battery were restricted to the Anagram and Snap tasks.

The research questions remain the same as for the neuropsychological tests in that the first issue dealt with initial levels of performance on the cognitive tasks, and the second dealt with changes in performance across testing sessions. In terms of levels of performance, each stroke patient's scores are presented in both tabular and graphical format, as was the case in Chapter 7. Initial levels of performance are reported in terms of z-scores where the normative sample is the comparison group. Raw scores on each of the measures are presented in graphical format where the scores of the normative sample and the age-matched controls are also presented. As indicated in Chapter 5, there were no significant age differences among older and younger members of the normative sample on either the Snap or Anagram tasks, and absence of these effects are readily apparent in the Figures. They are presented to be consistent with the approach taken in Chapter 7.

The cognitive tasks differed from the neuropsychological tasks in that dual task conditions were incorporated into the task in latter testing sessions, with the assumption that they would induce a decrement in performance. This change in task requirements meant that, unlike the neuropsychological tests, the use of regression slopes were no longer a viable measure of learning. Thus, three indicators of learning were adopted for the cognitive tests. The first involved the temporal stability measure that was used in Chapter 7. This measure again allowed for the

determination of significant deviations from initial performance. In this instance a significant improvement could be expected with repetition of the test, but a significant deterioration in performance with the addition of a dual task component might also be expected. The adoption of dual task components to the cognitive tasks represents the second indicator of learning. The normative group, at least in the two-event condition, was characterised by deterioration in performance on the session when the taxing secondary task was administered. This was followed by improved performance on the session immediately following. Participants rapidly adapted to the changes in dual task requirements. This pattern of dual task effects serves as a marker for evaluating the patients with stroke' performance on these tasks.

The third indicator of learning involves the somewhat unexpected finding in the normative sample, that word length effects in the Anagram task were quite pronounced in the early testing sessions, with performance on the 5-letter anagrams being particularly slow. However, in the latter testing sessions, the word length effect was severely attenuated. To the extent that the patients with stroke' performance on the task mirrors that of the normative sample, the attenuation of the word length effect would also be expected.

As was the case in Chapter 7 the performance of GL, KS, MF and LL are presented in a case study format. TB was not able to complete either of the cognitive tasks because of his vision problems. Essential methodological information that was presented in Chapter 7 is briefly repeated again purely as an aide for the reader. The results section examines initial performance of the patient, the extent to which they show dual task effects, and any attenuation of the word length effect.

## **8.2 Case Study 1: GL**

### **8.2.1. Participant Background**

GL is an 82-year-old male admitted to hospital with a subdural hematoma.

## **8.3 Method**

### **8.3.1 Test Administered**

The cognitive tests that were administered on each session are again summarised in Table 8.1.

Table 8.1

*Cognitive Tests Administered to GL across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Snap</i>	Single Task	✓	✓	✓			
	Dual Task						
	Single Event				✓		
	Two Events					✓	✓
<i>Anagram</i>	Single Task	✓	✓	✓			
	Dual Task						
	Single Event				✓		
	Two Events					✓	✓

## 8.4 Results

### 8.4.1 Initial Levels of Performance and Temporal Stability

GL's performance on the cognitive tasks are summarised in Table 8.2 and Figure 8.1.

On initial testing GL was within normal bounds on both the Snap and Anagram tasks. Moreover, GL was never in the clinical zone on any of the six testing occasions on the Snap task. The temporal stability data showed a similar pattern. He consistently improved over the first four sessions such that he was outside the 95% confidence levels at Sessions 3 and 4. With the introduction of the two-event dual task condition, performance returned to within the 95% confidence limits, but was again outside the confidence band on Session 6.

On the Anagram task, GL was in the normal range on initial test, but deteriorated on the next two sessions where his performance was in the clinical range. On Session 4, performance was significantly better than initial performance as was the case on Session 6, where performance was within normal range functioning. The "blip" at Session 5 can be attributed to the introduction of the two-event dual task condition.

Table 8.2

*GL's Performance on the Cognitive Tasks, Expressed in Z-scores, across Test Sessions*

	Test Sessions					
	1	2	3	4	5	6
Snap	0.60	1.63	1.19	0.97	0.96	0.93
Anagram	1.76	3.33	4.19	3.02	2.79	1.26

#### **8.4.2 Dual Task Performance.**

In both Snap and Anagram tasks, the single-event detection dual task condition was introduced for a single session during Session 4. The addition of the second task had minimal noticeable effect on performance. There was no effect on primary task performance. Completion time on both the Snap and Anagram tasks were significantly better than baseline at Session 1, there was no adverse effect on pair detection in the Snap task, and no marked changes in anagram errors. The only signs of a dual task effect were in the event detection secondary task and this was limited to the Snap task where a small number of events were not detected. In the Anagram task, event detection was perfect.

The two-event dual task condition was implemented in Sessions 5 and 6. Here there was strong evidence for dual task effects in the fifth session with performance on the sixth session showing adaptation to the changed nature of the task. In the Snap task there was a slowing on the completion time such that performance returned to within the 95% confidence band. Moreover, one of the pairs was not detected and two of the event cards were not detected. In the sixth session completion time was again faster than baseline, all pairs were detected as were all the event cards.



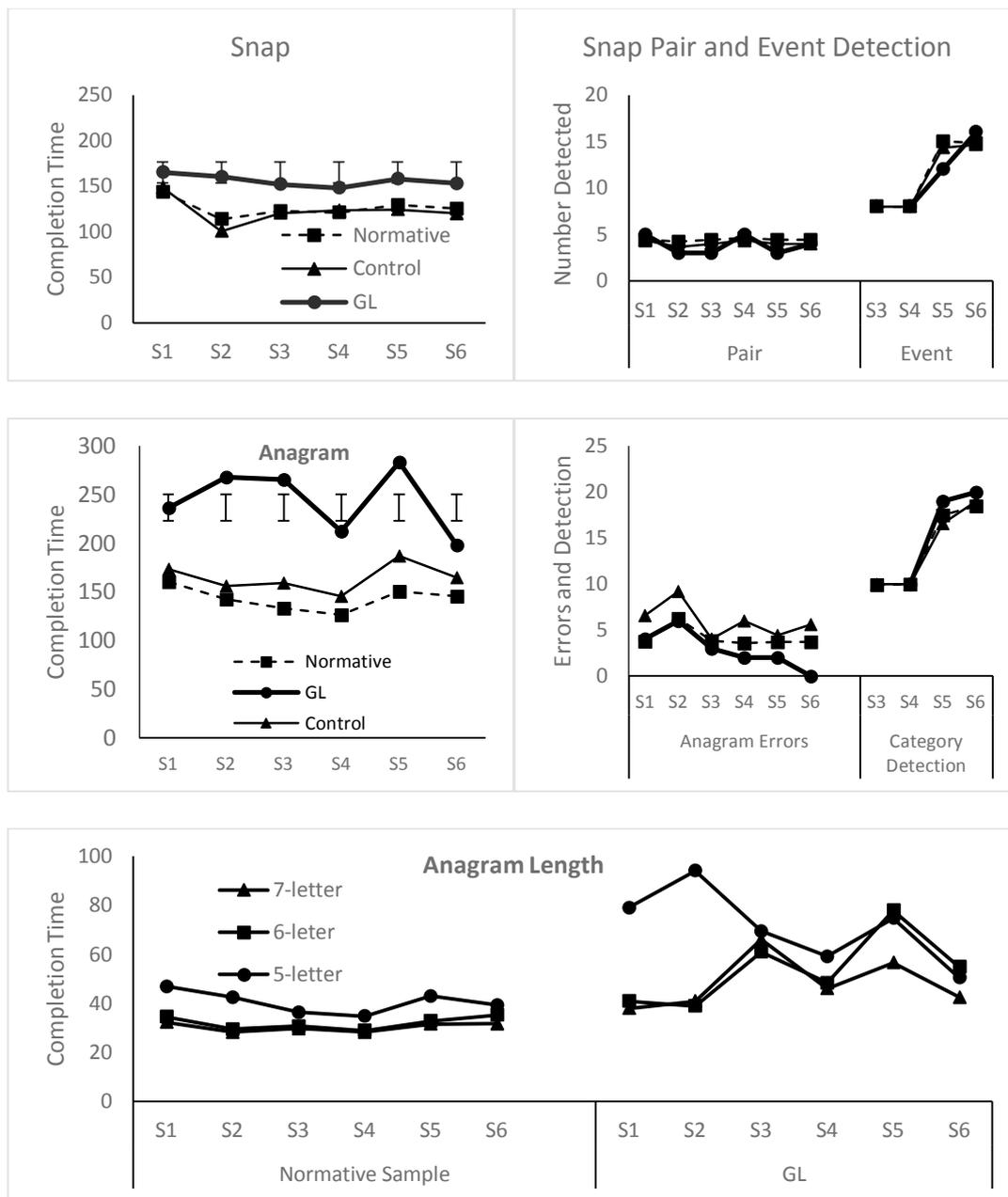


Figure 8.1. GL's performance on snap and Anagram tasks as a function of testing session.

In the Anagram task, the clearest effect of introducing the two-event condition was a marked slowing in completion time in the fifth session, but the same pattern was found in the third session with the 6-letter and 7-letter anagrams. In both instances, there was marked improvement in the following session. That is, GL quickly adapted to the dual task requirements. In the final session, completion time improved to the point that even under high difficulty dual task conditions, performance had improved above baseline levels both in terms of completion time, a reduction in anagram solution failures and in terms of event detection.

In short, dual task effects were observed only when two events had to be detected, and this effect was only observed on a single session. The dual task effects were obvious only in completion times on the primary task. GL quickly adapted to the two-event dual task condition in that all measures showed improvement in the final session when compared to the previous session and to baseline.

#### **8.4.3 Letter Length in the Anagram Task.**

Figure 8.1 shows the changes in anagram solution times as a function of letter length. GL's pattern of performance on this task was very similar to that found in the normative sample. As was the case in the normal performance there were large differences in solution time between 5-letter and 6- and 7-letter words in the initial sessions. With additional exposure to this task, there was marked improvement in the solution times for 5-letter words such that by the later sessions the letter length effect had been severely attenuated. This outcome indicates that on this task, normal learning outcomes were observed even if overall solution times were somewhat slower than those of the normative sample.

### **8.5 Discussion and Overall Evaluation**

In the case of the Snap task, performance was within normal limits on all sessions for both completion time and pair detection. In addition there was evidence that performance improved on the sessions where it was expected. Moreover, dual task effects paralleled those of the normative sample. To the extent that both quantitative and qualitative aspects of this task are within normal limits, it would appear that the stroke has not had a massive impact on this task and as such working memory updating as reflected in pair detection and completion time was not impaired.

GL was also within normal limit on initial testing on the Anagram task. While there were sessions where completion time fell into the clinically impaired range, there was evidence of learning across sessions in terms of completion time and in reduced numbers of errors. Dual task performance was qualitatively similar to the normative sample in completion time and quantitatively similar in event detection.

## 8.6 Case Study 2: MF

### 8.6.1 Participant Background

MF is an 82-year-old female retired teacher who was admitted to hospital with a diagnosis of a right frontal lobe subacute infarct.

## 8.7 Method

### 8.7.1 Test Administered

The cognitive tests that were administered on each session are again summarised in Table 8.3.

Table 8.3

*Cognitive Tests Administered to MF across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Snap</i>	Single Task	✓	✓				
	Dual Task						
	Single Event				✓		
	Two Events						
<i>Anagram</i>	Single Task	✓	✓				
	Dual Task						
	Single Event			✓	✓		
	Two Events					✓	

### 8.7.2 Test Notes

The Snap task was the last test on the battery and on Session 3 MF was required to go for physiotherapy before this task could be completed. As noted in Chapter 7, MF indicated that she was fatigued and decided not to complete the last test. She could not be tested again before discharge.

Due to insufficient data, her performance on the Snap task is not reported.

### 8.8 Results

#### 8.8.1 Initial Levels of Performance and Temporal Stability

MF's performance on the cognitive tasks are summarised in Table 8.4 and Figure 8.2.

MF was in the clinical range on the first test session. She remained in the clinical range on the next three sessions but was in the normal range on the final session. Strong and consistent improvement were evident in terms of temporal stability confidence intervals. She was significantly better than baseline performance by the second session. Inspection of Figure 8.2 indicates that her improvement across sessions was substantially greater than was the case with either the normative sample or the age-matched controls, albeit that she was coming from a slow initial position.

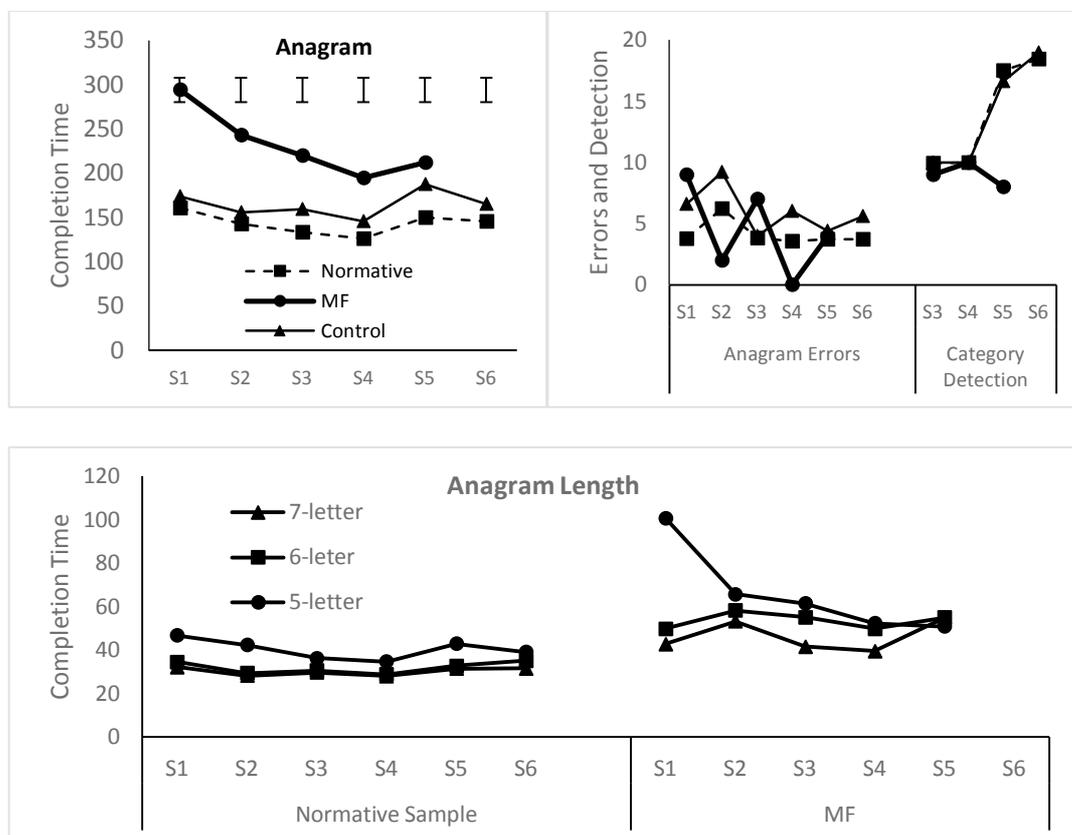


Figure 8.2. MF's performance on the Anagram task as a function of testing session.

Table 8.4.

*MF's Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions*

	Test Sessions					
	1	2	3	4	5	6
Anagram	<b>3.10</b>	2.67	2.74	2.39	1.29	

### 8.8.2 Dual Task Performance

In the Anagram task, single-event detection was implemented in Sessions 3 and 4. The addition of the second task had a moderate effect on performance. Completion time on the Anagram task became consistently faster, with no evidence of interference by the dual task. There was evidence of dual task effects on anagram errors. In Session 3 there was an upturn in errors but MF quickly adapted, producing a perfect anagram score in Session 4. MF showed no dual task effects in event detection, obtaining a perfect score in Session 4.

The two-event dual task condition was implemented for a single session in Session 5. There was strong evidence for dual task effects in this session, with completion time, number of anagram errors and category detection all being affected. Completion time markedly increased during Session 5, but by this stage her absolute levels of performance were within the normal range. Number of anagram errors increased from Session 4, however this increase was within the 95% confidence interval band. Performance on category detection was affected, with performance changing from perfect at Session 4 to less than 50% accuracy in Session 5.

In short, dual task effects were observed for both single event and two event tasks. The single event effects were reflected only in an increase in anagram errors produced. The two event effects were reflected in an increase in completion time, anagram errors and a significant decrease in category detection. MF quickly adapted to the single-event dual task condition in that Session 4 showed substantial improvement compared to the previous session and compared to baseline.

### **8.8.3 Letter Length in the Anagram Task**

Figure 8.2 shows the changes in anagram solution times as a function of letter length. MF's pattern of performance on this task was very similar to that found in the normative sample, and with the exception of Session 1, the magnitude of the effect was quite similar to that of the normative group. There were large differences in solution between 5-letter and 6- and 7-letter words in the initial session. In the second session, however, there was marked improvement in the solution times for the 5-letter words. In the final three sessions, improvement was roughly equivalent for the three lengths, but in these sessions the letter length effect had been substantially attenuated. This outcome indicates that on this task, normal learning outcomes were observed even if overall solution times were a little slower than those of the normative sample.

## **8.9 Discussion and Overall Evaluation**

All the indicators are that MF showed normal learning on this task. There was constant and substantial improvement across sessions, such that by the final session her scores were in the normal range. Performance on the dual task components of the Anagram task suggest that MF was able to handle the one-event dual task situation quite easily in terms of there being no appreciable cognitive load effect on completion time in either Session 3 or Session 4, and performance on the detection task was near perfect. In Session 5 where the two-event prospective memory task was included, there was a slowdown in completion time consistent with increased cognitive load, but performance on the detection task was again error-free. Finally, strong letter-length effects on the initial test session were severely attenuated on the final three sessions, as was the case in the normative sample. In short, there is clear evidence of cognitive recovery in this task.

## **8.10 Case Study 3: KS**

### **8.10.1 Participant Background**

**KS** is a 55-year-old male, farmer admitted to hospital following a Middle Cerebral Artery stroke in his right hemisphere.

## 8.11 Method

### 8.11.1 Test Administered

The cognitive tests that were administered on each session are again summarised in Table 8.5.

Table 8.5

*Cognitive Tests Administered to KS across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Snap</i>	Single Task	✓					
	Dual Task						
<i>Anagram</i>	Single Event						
	Two Events						
	Single Task	✓	✓				
	Dual Task						
	Single Event			✓	✓		
	Two Events					✓	

### 8.11.2 Test Notes

As mentioned in Chapter 7 KS had general concerns regarding the utility of cognitive rehabilitation and specific aversion to the Snap task, due in part to the instructions that referred to a children's card game. KS indicated that he "felt like he was in preschool" when doing the task. In light of his antipathy to the task, it was not administered on subsequent sessions.

## 8.12 Results

### 8.12.1 Initial Levels of Performance and Temporal Stability

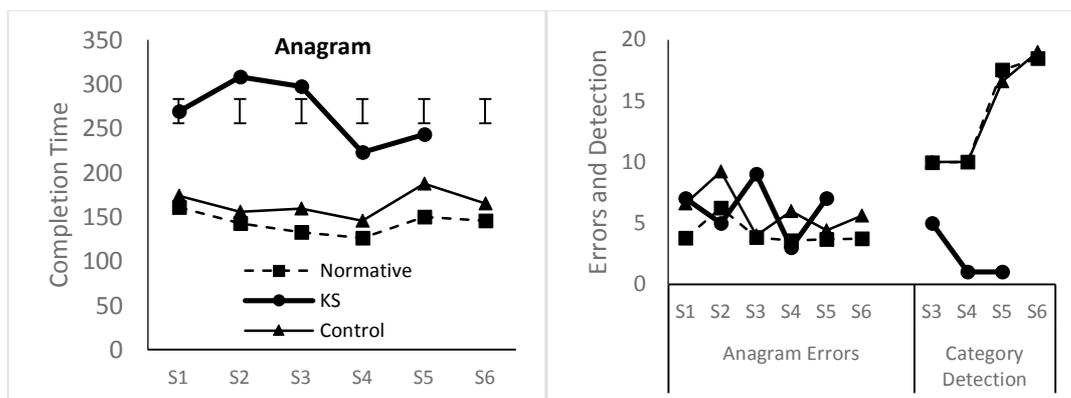
KS's performance on the cognitive tasks are summarised in Table 8.6 and Figure 8.3.

KS was in the clinical range on the first test session. He remained in the clinical range on the next three sessions but was in the normal range on the final session in which he participated. In terms of temporal stability, KS deteriorated on the second session and was again significantly worse than baseline on Session 3 when the single-task dual task condition was first introduced. However, there was a dramatic improvement from Session 3 to Session 4. While there was a slowing on Session 5 with the introduction of the two-event dual task condition, performance on Session 5 was still significantly better than baseline performance.

Table 8.6

*KS's Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions*

	Test Sessions					
	1	2	3	4	5	6
Anagram	2.52	4.40	5.19	3.38	1.94	



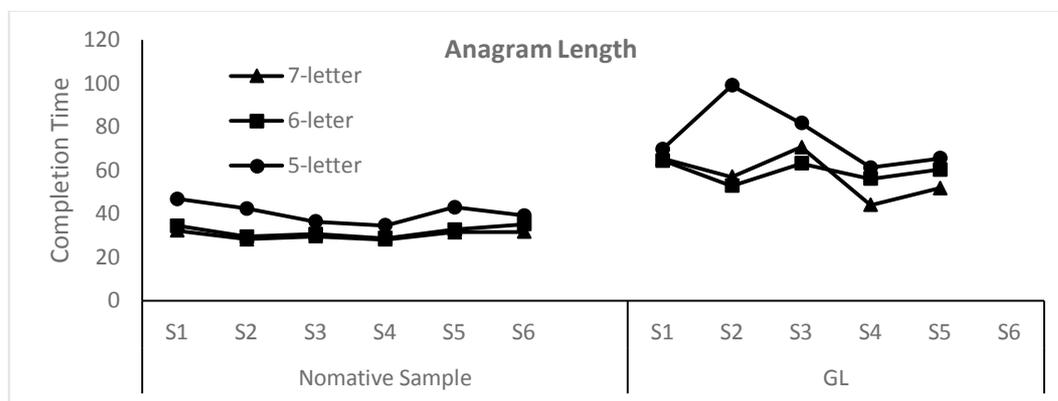


Figure 8.3. KS's performance on the Anagram task as a function of testing session.

### 8.12.2 Dual Task Performance

The evidence would suggest that KS found the Anagram task to be difficult in many respects. His initial performance was in the clinical range, and in the second and third sessions completion time deteriorated. In Session 3 when the single-event secondary task was introduced for the first time, the expected slowing can be observed in the 6- and 7-letter anagrams with substantial improvement in the 5-letter anagrams. However, while the completion times showed the effects of the implementation of a secondary task, event detection was well below normative levels on all sessions, with only half the events successfully detected in most instances. In the fifth session, KS identified the first instance of the nominated category but no subsequent category members were identified. It is not clear whether this was a failure to remember the change in dual task instructions or whether the primary task was so difficult that all cognitive resources were devoted to the primary task. In either case it is clear that KS did not handle the dual task requirements well in any of the sessions.

### 8.12.3 Letter Length in the Anagram Task

Letter length effects could almost be described as identical to that found with the normative sample. Performance on the 7- and 6- letter words showed the same pattern as the normative sample.

in that there was improved performance when there was no change in task requirements from the preceding sessions (Sessions 2 and 4), and there was a slowing in completion time when there was a change to dual task requirements (Sessions 3 and 5). Furthermore, the completion time advantage for the longer anagrams over the 5-letter anagrams was present in the early sessions and was attenuated in the latter sessions as was the case in the normative sample. In fact, KS departs from the normative pattern only on the first two sessions with the 5-letter anagrams. It is not clear whether Session 1 was abnormally fast, or Session 2 was abnormally slow, or both.

### **8.13 Discussion and Overall Evaluation**

There were a number of aspects to KS's performance that were indicative of slower, but normal performance. These indicators were reflected in performance on the primary task. Inspection of the temporal stability confidence intervals indicated that by the later testing sessions, completion times had improved significantly compare to initial test scores. In addition, there was clear evidence of dual task slowing when both single- and two-task conditions were introduced. Likewise, the quick adaptation to dual task requirements was observed in Session 4. Finally, letter length effects were attenuated in the later testing sessions as per the normative group. Thus, there was substantial evidence to suggest that KS showed normal learning in terms of completion time.

The problematic aspect for concluding normal recovery of task function stems from performance on the secondary task. KS was particularly poor at the event-detection task, and this was particularly marked in the two-event detection component where KS failed to report instances from the second nominated category. One possible explanation is that the pattern reflects a speed-accuracy trade-off. That is, KS devoted his cognitive resources to performing the primary task at the expense of the secondary task. Whatever, the explanation it is clear that KS's performance on this task shows only some of the characteristics of normal performance.

### **8.14 Case Study 4: LL**

#### **8.14.1 Participant Background**

LL is a 79-year-old male admitted to hospital with a stroke impacting his left hemisphere.

## 8.15 Method

### 8.15.1 Test Administered

The cognitive tests that were administered on each session are again summarised in Table 8.7.

Table 8.7

*Cognitive Tests Administered to LL across Six Test Occasions.*

Task	Condition	Tests					
		1	2	3	4	5	6
<i>Snap</i>	Single Task	✓					
	Dual Task						
<i>Anagram</i>	Single Event						
	Two Events						
	Single Task	✓	✓				
	Dual Task						
	Single Event			✓	✓		
	Two Events					✓	✓

### 8.15.2 Test Notes

The Snap task was administered as the final test in the battery in the first session. LL experienced particular difficulty understanding the snap instructions. Numerous methods of explanation and demonstration were used. Moreover, the repetitive nature of the Snap task made attending and concentrating difficult such that reliable data were not able to be gathered. Consequently, the Snap task was not administered on subsequent sessions.

8.16 Results

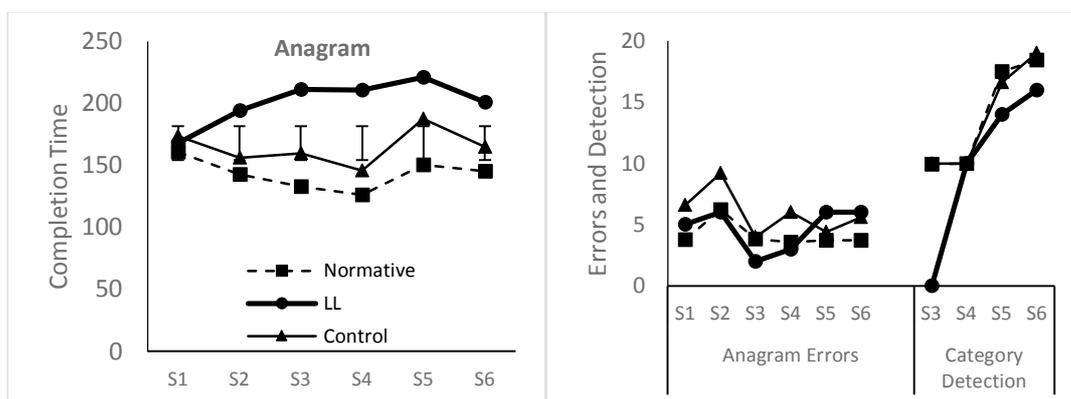
8.16.1 Initial Levels of Performance and Temporal Stability

LL’s completion time performance on the Anagram task is summarised in Table 8.8 and Figure 8.4. LL was in the normal range on the first test session. However, temporal stability measures show that completion time deteriorated from baseline on Session 2 and remained at below baseline measures on all subsequent sessions. The deterioration was such that on Sessions 3 and 4, performance was in the clinical range. However, in Sessions 5 and 6, performance was once again in the normal range. Solution errors were no more pronounced than for either the normative or control groups.

Table 8.8

*LL’s Performance on the Anagram Task, Expressed in Z-scores, across Test Sessions*

Test Sessions						
	1	2	3	4	5	6
Anagram	0.17	1.37	2.48	2.96	1.48	1.32



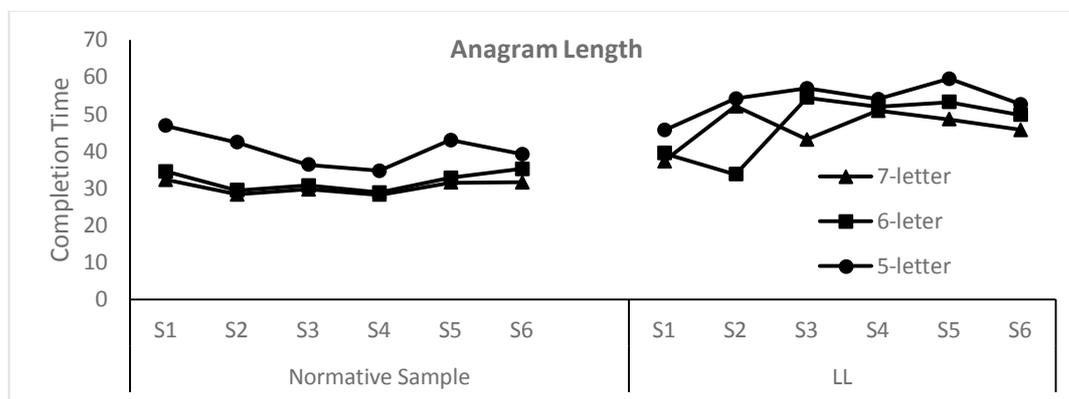


Figure 8.4. LL's performance on the Anagram task as a function of testing session.

### 8.16.2 Dual Task Performance

Evaluating dual task effects was made difficult by performance over the first three sessions. In contrast to the normative sample, LL showed substantial slowing over the first three sessions. Notable slowing in completion time between Session 2 and Session 3 was consistent with the implementation of the secondary single-event detection task in Session 3. However, there were no corresponding changes in error rates on anagram solutions, and there were no detections of the nominated category. That is, while the instructions introduced the dual task component in Session 3, there was no evidence that LL processed or maintained these task changes during the session. It would seem that dual task instructions were only processed and acted on in Session 4. In Session 4, there was no slowing compared to the previous session, error rates were within normal limits, and single-event detection was at the 100% level. When the two-event dual task condition was introduced in Session 5 all the components of the dual task effect were present. There was a slowing in completion time, an increase in solution failures, and event detection was prone to errors. In this case, items from the first nominated category were correctly identified but errors on the second category dominated. In the final session, completion time decreased and event detection increased, suggesting that LL adapted to the dual task quite well in this session.

### 8.16.3 Letter Length in the Anagram Task

Figure 8.4 shows the changes in anagram solution times as a function of letter length. LL's pattern of performance on this task is hard to interpret because of the increase in solution times across the first three sessions. Performance on the 7-letter anagrams on Session 2 was particularly slow for some reason. If this measure was in line with the other 7-letter measures, then performance would be largely consistent with the normative sample. As was the case in the normative sample, length effects were attenuated in the later sessions, suggesting that normal learning occurred across the last three sessions.

### **8.17 Discussion and Overall Evaluation**

LL's performance on the Anagram task appears to have changed in two phases. He was well within normal limits on initial testing. However, on Sessions 2 and 3 there was a deterioration in terms of completion times, such that by the third session LL's performance was in the clinical range. While there was deterioration in completion time there was no corresponding change in solution errors. LL did not appear to process the introduction of the dual task component in the third session.

The second phase, Sessions 4 to 6, shows the hallmarks of normal performance in that performance returned to within normal limits, dual task effects were detectable in terms of changes in completion time and event detection, and the attenuation of letter length in solution time was apparent.

### **8.18 Comparing Patients**

In Chapter 7 when creating composite scaled scores it was possible to distinguish three different profiles that reflected three different learning outcomes. That is, normal practice effects could be distinguished from recovery, and both could be distinguished from a generalised learning deficit. KS showed both practice effects and recovery, MF showed practice effects but not recovery, and the remaining patients showed a generalised learning deficit. In Figure 8.5 the data of all four patients has been presented using the same format as that used in Chapter 7. That is, the raw data from the primary measure has been converted to scaled scores, and the dark region at the bottom of

each Figure indicate clinical levels of cognitive impairment. The Figure also presents regions where the dual-task has been introduced. The darker vertical region reflects where there has been a change in dual-task requirements (Sessions 2-3, and 4-5) and the lighter vertical region reflects where the dual-task requirements remain the same across sessions (Session 3-4, and 5-6). The negative sloping line in the darker regions represent dual task effects, and adaptation to dual task requirements are represented by a positive sloping line in the lighter regions.

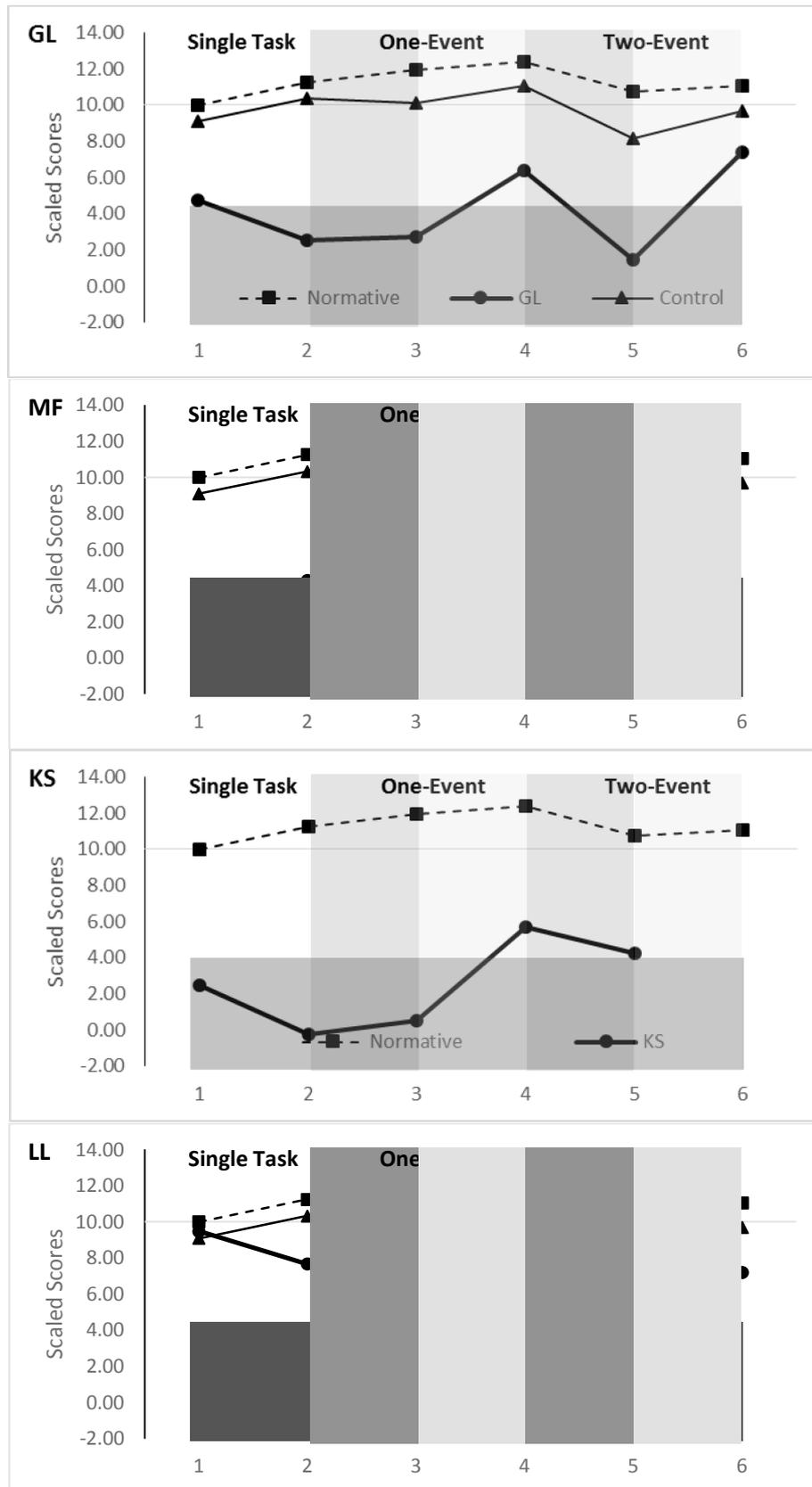


Figure 8.5. Patients with stroke' performance on the Anagram task expressed in scaled scores as a function of dual task condition and testing session.

The outcomes using the Anagram task are quite distinct from those using the cognitive tests. In the neuropsychological tests three distinct profiles were observed where only one participant, KS, showed evidence of natural recovery. As Figure 7.9 indicated, even though there was rapid improvement in performance, performance was still in the clinical region on his final session. As Figure 8.5 suggests that while profile across the early sessions differ between patients, all four show the same profile in Sessions 4 to 6. For GL, KS and LL the transition from Sessions 1 to 2 and from Sessions 2 to 3 can be characterised as an initial deterioration in performance followed by stable performance. As for the normative sample, the introduction of a single-event dual task condition did not overly impact on performance for these participants. For MF, performance increased substantially across all sessions, which is indicative of recovery. The transition from Sessions 3 to 4 was characterised by large improvements for GL, KS and MF such that all improved such that by Session 4 all patients were in the non-clinical range. The transition from Session 4 to 5 corresponds to the introduction of the two-event dual task condition, and all four patients with stroke showed a dual task decrement, as did the normative and control groups. Even with this added difficulty, MF and KS produced levels of performance that were within normal range functioning. Performance rebounded in Session 6 for GL, LL and the normative and control groups, and GL's performance had again returned to within normal limits.

In short, all participants were within normal range functioning at Session 4 and all were within normal limits on their final session. All showed dual task decrements when dual task difficulty increased to having to detect two different stimuli. Where it was possible to observe, rapid adaption to changes in dual task requirements were also observed in all patients.

### 8.19 General Discussion.

This chapter, like the preceding one, explored the issues associated with the repeated administration of cognitive tests to a group of patients with stroke. Interest centred on the degree of initial impairment and the detection of improved performance across subsequent test sessions.

Dual task effects were examined with reference to three aspects of performance: completion times, the number of errors on the primary task, and correct detection of the nominated event on the secondary task. Slowing in completion time or an increase in the number of anagram errors, or both were the markers of dual task effects on the primary task. In the single-event dual task condition, dual task effects were not always evident in the completion time data (MF), and in some instances were more obvious in the easy conditions (7-letter and 6-letter) than in the total score. However, in such instances, dual task effects were present in the error data. The more frequent outcome was the presence of both indicators. The outcomes were less variable in the two-event condition. The dual task decrement in total completion time was evident in all patients. Performance was more variable on the secondary task. Two of the patients handled both the primary and secondary tasks reasonably well but others found the secondary requirement particularly challenging.

One somewhat unexpected but robust outcome was the rapid adaptation to the changes in dual task requirements. In the sessions immediately following a session where a dual task decrement had appeared, there was a substantial improvement in all aspects of performance. This was evident in Session 4 and in those who were tested in Session 6. Thus, the last few sessions can be characterised by a dual task decrement being observed whenever secondary requirements changed followed by a recovery of performance in the subsequent session. This pattern provides another indicator that the patients with stroke were showing enhanced recovery on this task.

The final aspect of performance involved anagram length effects on completion time. Normative performance on this task involved letter length effects in the early testing sessions but an attenuation of these effects with increasing exposure to the task. The performance of the patients with stroke was more variable on the early sessions, but by and large the expected letter length effect

was present in the early sessions as they were in the normative sample. Performance was less variable in the latter sessions with all patients with stroke showing attenuated length effects.

In sum, dual task effects were present for all patients, and these were most clearly seen with the introduction of the two-event dual task condition. More importantly, all patients with stroke were within normal range functioning by Session 4, and this was maintained in their final session. For these patients, improvement on this task was more pronounced than for the normative group. Thus, changes in absolute levels of performance and dual task effects on the Anagram task suggest that the patients with stroke were performing this task without substantial impairment by the end of testing. More importantly, the results are consistent with the working memory training literature where improvements in performance are greater when the cognitive load is systematically increased during training, rather than when a task is simply repeated across training sessions (Holmes et al., 2009; Klingberg et al., 2002, 2005). In short, the comparative improvement on the Anagram task, relative to the other tests, indicates that tasks that manipulate task difficulty are more potent in facilitating recovery than repeated practice alone.

## **Chapter 9: Evaluation of Project**

### **9.1 Chapter Overview**

This chapter has four aims. The first is to briefly review the key issues derived from the literature that formed the rationale for the studies and informed the specific aims of the research. The second is to recapitulate what was done and what was found. The third is to critically evaluate the outcomes and their theoretical implications with respect to rehabilitation practice during the acute phase of stroke. The fourth is to comment on the limits of the current study and identify areas for future study.

### **9.2 Rationale for the Study**

Suffering a stroke leads to the destruction of cells in the brain around the area where bleeding or obstruction has occurred. However, the effects can be more widespread, affecting other undamaged areas of the ipsilateral hemisphere that are networked to the damaged area, and can extend to processes occurring in the contralateral hemisphere. The current thesis is built around two sets of related changes that occur in the aftermath of the stroke. The first is that biological recovery mechanisms in the brain come into play to ameliorate the damage either by restoring existing functioning or by compensating for the damage via neural plasticity changes (Wieloch & Nikolich, 2006). The second set of changes involves the behavioural recovery of functions that occurs during the same time frame (Cassidy, Lewis, & Gray, 1998; Duncan et al. 1992; Kertesz, 1988). These biological and concomitant behavioural changes are known as natural recovery. The main theme of the thesis concerns natural recovery of cognitive behaviours, and whether or not such behavioural changes can be observed and facilitated during the acute phase of stroke.

The literature review indicated that the acute phase of stroke could be characterised in terms of patients having multiple impairments, with many of these impairments being resolved at 3 months when comprehensive batteries of neuropsychological tests were used to assess cognitive functioning

(Hurford et al., 2013; Nys et al., 2007; Tatemichi et al., 1994). Moreover, the number and profile of impairments differed from patient to patient such that no single stroke profile existed.

While cognitive impairments are readily detected in the acute phase, there is no published research that directly addresses cognitive changes during the first 3 months following the stroke. In studies assessing performance after the three-month point, the research shows that subsequent recovery can often take place (Desmond et al., 1996, Tham et al., 2002) but performance gains are slow and are small in magnitude (Hochstenbach et al., 2003; Sachdev, 2004b). Given the absence of literature that explores the recovery process during the acute phase itself, the first aim of the current research was to explore the recovery process during the acute phase of stroke. This was done by examining improvements in performance, or lack thereof, across repeated administration of a series of standard neuropsychological tests.

Given that cognitive recovery occurs, it is surprising that there are very few published studies that have explored interventions in the cognitive areas of interest. Of the few that have been published, all interventions have been conducted during the chronic phase of recovery (Stablum et al., 2000; Vallat et al., 2005; Westberg et al., 2007; Nordvik et al., 2012). The outcomes of those studies are reasonably encouraging to the extent that some studies show improvement in performance on the training task across training sessions (Nordvik et al., 2012), some studies show near transfer effects (Stablum et al., 2000; Westberg et al., 2007), and one study shows far transfer effects (Stablum et al., 2000).

These training studies stand in contrast to the studies that have explored recovery in the chronic phase. In the training studies, substantial improvement in performance has been observed whereas natural recovery studies show weak learning effects when assessed through neuropsychological tests. One possible explanation for such differences is the type of task used to measure performance, with cognitive training tasks being more potent in assessing/facilitating behaviour change. To the best of the author's knowledge, no studies have been published where interventions have started during the acute phase of recovery, so it is not possible to evaluate

recovery versus training effects, or neuropsychological tests versus cognitive tests on the basis of literature alone. Consequently, the second aim of the thesis was to explore the changes in performance associated with repeated exposure to a group of standardised neuropsychological tests, and a group of tests that have their roots in the theory and practice of cognitive psychology.

### 9.3 Outcomes

#### 9.3.1 Neuropsychological Tests

The expectations concerning repeated testing critically depended on whether or not the stroke had a detrimental impact on the cognitive domain under consideration. Where initial testing indicated that performance was not impaired, learning effects in the stroke patient were expected to be similar to those found in the normative sample (Nordvik et al., 2012). Recovery only became relevant when initial testing indicated that there was a cognitive impairment on the test. Five of the measures (digit span, Stroop, phonemic and semantic fluency and the Rey Tangled Lines test) were directly relevant to examining recovery in that these tasks were repeated without any modification across all sessions.

In addressing the presence of recovery, a number of aspects of performance were considered. The first was the absolute level of performance on which the presence or absence of a cognitive deficit was defined. The focus was on evaluation of performance on the initial test session for a number reasons to determine whether: impairments could be detected, comparisons could be made to existing literature concerning the prevalence of impairments, and whether performance improved over baseline on subsequent sessions. A consistent improvement across sessions for measures where initial test performance indicated an impairment was taken as an indicator that recovery was occurring.

**9.3.1.1 Initial levels of impairment.** A number of findings emerged that are generally consistent with existing literature. All participants showed evidence of impairment on one or more of the tasks. Just as importantly, all participants showed normal range functioning in one or more tasks. In four of the five participants, the split of impaired versus intact performance was in the 50% range.

In the case of TB, there were greater levels of impairment than there was intact performance. In this sample there were no participants where impairment was restricted to a single task. In short, even with a small sample of only five participants, some showed deficits across a wide range of tasks and measures and others showed selective deficits. The results of the research are entirely consistent with what is currently known in terms of the variety and breadth of stroke's impact on cognitive performance in the immediate aftermath of a stroke (Jaillard et al., 2009; Nys et al., 2007; Tatemichi et al., 1994).

**9.3.1.2 Tasks affected by stroke.** In this sample, all participants were within normal limits at initial testing for both forward and backward Digit Span tasks, and subsequent testing sessions indicated that performance did not appear to be overly impacted by the stroke. The Rey Tangled Lines and colour-word naming component of the Stroop tasks represent measures where there were differential levels of impact. Thus, some participants showed no baseline impairment on these tests and some displayed either moderate or severe impairments on these tasks. Most importantly, the people who were impaired on one of these particular tests were not necessarily the same people who showed deficits on other tests.

The third group of tests were those that all patients experienced difficulty with on initial testing. All patients were in the clinical range on the read and colour naming components of the Stroop tasks, and the same was true of the three measures on the semantic fluency task. The patients remained in the clinical range throughout testing.

The current findings in many respects are similar to those in the studies reviewed earlier. In the majority of those studies, cognitive impairments due to stroke were examined at the domain level rather than at the test level. However, there are studies that do report performance at the test level, or where performance at the domain level can be compared to the current results. As indicated in Chapter 2, many of the studies show that impaired performance can be detected on most measures. For example, impairments were observed in digit span (Jaillard et al., 2009; Planton et al., 2012; Sachdev et al., 2004a), in verbal fluency (Jaillard et al., 2009; Maidureira et al., 2001; Planton et al.,

2012; Pustokhanova & Morozova, 2013; Sachdev et al., 2004a), and all measures of the Stroop task (Planton et al., 2012). As a result, some of the more interesting contrasts involve between-measure comparisons such as the impairment rates between forward and backward digit span, and between phonemic and semantic fluency.

The van Zandvoort et al. (2005) study is one of the most relevant because testing of 57 patients with stroke was done in the acute phase and data of task completion (difficulty) was reported, as well as initial levels of performance and follow-up data. Their battery included digit span (combined scores), phonemic and semantic fluency tasks. They also included the Trail Making Test that bears some similarities with the Rey Tangled Lines test in that both are visual tasks that require following a trail as quickly as possible. Digit span was the most frequently completed test with 90% of the participants being able to complete the task. Of those who completed it 53% were not impaired and 29% were severely impaired. The current data appear to be consistent with this in terms of being the one test where there was least initial (and subsequent) impairment. When looking at the differences between forward and backward recall, performance deficits were more apparent on backward span than on forward span in the current studies, an outcome that is also found in other studies (Jaillard et al., 2009).

For the verbal fluency measures 66% of participants did the task in the van Zandvoort et al. (2005) study. There were clear differences in outcomes between phonemic and semantic tasks in that 52% of patients had severe impairments on the semantic task and only 23% of patients had severe impairments on the phonemic task. This pattern of the semantic task being more sensitive to stroke than the phonemic task is also present in the current data and in other studies as well (Jaillard et al., 2009; Maidureira et al., 2001; Pustokhanova & Morozova, 2013).

Finally, in the van Zandvoort et al. (2005) study, the Trail Making Test was difficult in that only 50% of participants did the task, and of those who did, over 50% showed severe impairment on both Part A and Part B of the test. In the current data, the Rey Tangled Lines was the test that

showed the most extreme scores in terms of z-scores. Both sources suggest that complex visual processing is particularly susceptible to stroke.

**9.3.1.3 Learning.** Since this is the first study to examine the learning that goes on with repeated testing during the acute phase, there is no background literature with which to compare current outcomes. Expectations were based on the logic that if a cognitive function was unimpaired in a stroke patient, they should show a similar learning trajectory to the normative sample (Nordvik et al., 2012). This assumption has two implications at the individual task level. For a start, all tasks that were initially unimpaired should show the same pattern of learning as the normative sample. An unexpected result would be one where the normative sample showed learning on all tasks, but a stroke patient showed learning on some of the tasks but no learning on others. The second assumption is that expectations about whether or not performance on an individual task would be expected in the stroke are determined by normative performance on that task. Thus, not all tasks need show learning effects, but if the stroke patient's initial performance is in the unimpaired range it should reflect the same degree of learning as per the normative sample.

Observing recovery in impaired tasks suffered the same interpretive problems. In the case of digit span and Rey Tangled Lines, improvement in the tasks would be indicative of recovery. However, in the case where the normative sample showed repetition effects, the ability to distinguish between normal repetition effects and recovery of function becomes problematic. Desmond et al. (1996) recognised this problem and operationalised recovery in terms of a person having to show improvement that was 2 standard deviations above that exhibited by a control group. The patients with stroke had to show accelerated learning compared to the control group.

The current normative sample showed strong learning effects in all measures save for forward digit span and Rey Tangled Lines. Thus, the absence of learning on digit span and Rey Tangled Lines in the stroke group is not necessarily an indicator of abnormal performance in the stroke group because many of the normative sample did not show sustained learning in these tasks.

In looking at performance on all tasks, there was a consistent pattern across tasks for each participant. Either all the tasks showed the relevant learning effects or they consistently showed the absence of learning. This meant that it was feasible to calculate composite scores where raw scores could be converted to scaled scores and averaged.

Improvements or lack of improvement could be detected at the individual test level. All five patients with stroke demonstrated improved performance on at least one measure, and all showed no improvement on at least one measure. However, the overall outcomes were clearest when composite scores were generated for initially impaired and initially unimpaired tasks. Three quite distinct profiles were recognisable.

KS improved on the tasks that were initially unimpaired, as per expectation, and the rate of improvement was the same as for the normative sample. Thus, KS showed normal practice effects. More importantly, KS also improved on the tasks that were initially in the impaired range. Performance on these tasks improved at a much faster rate than that produced by the normative sample. KS showed accelerated learning in these tasks that is consistent with recovery of impaired function. For KS, therefore, there was evidence of substantial natural recovery in all affected domains over the testing period.

The second profile was that represented by MF's performance. On the initially unimpaired measures, she showed evidence of the improvement that was expected on intact cognitive domains. Like KS, the improvement in this task was at the same rate as in the normative sample. Again, normal practice effects were apparent. However, on the initially impaired measures, unlike KS there was no evidence of widespread and sustained improvement. There was no widespread evidence for recovery in any of the impaired domains.

The third profile was represented by GL, TB and LL, and was characterised by the absence of learning in both intact and impaired tasks. The absence of learning in the impaired domains is indicative of minimal recovery over the testing period, and as such this outcome was to some extent expected. The more problematic findings were that there was no improvement in the domains that

were initially intact. This suggests that one of the outcomes of the stroke for these patients is that they have sustained a generalised learning impairment. If so, this general learning impairment is similar to deficits in speed of processing, in that the deficit is recognised not on the basis of a single task, but is determined by a consistent pattern of performance across tasks (Hochstenbach et al., 1998; Planton et al., 2012). An alternative explanation is that in this early stage post-stroke, the level of “illness” is having a detrimental impact on all cognitive activity. In other words, they are simply too sick to perform well (van Zandvoort et al., 2005).

These three distinct profiles provide evidence that the tasks in the battery are sensitive to trauma caused by stroke not only in terms of detecting deficits in absolute levels of performance, but also in detecting two distinct types of learning effects. Learning in non-impaired measures (KS and MF) show that the tasks are also sensitive enough to detect improvement in the patients with stroke in the same way and at the same rate that improvement could be detected in the normative sample (Experiment 5.1). Most importantly, the tests were sensitive to natural recovery as shown by KS’s accelerated improvement in all areas in which he had been initially impaired. As such, the question as to whether or not natural recovery can be observed using the current test battery can be answered in the affirmative.

The profiles make one other point. While recovery during the acute phase of stroke can be observed, most of the participants did not show recovery during the testing. Thus, while the results show that natural recovery can be observed in a 2- to 3-week period in the acute phase of stroke, the more likely response is that such recovery will not be observed. This begs the question as to whether or not the lack of recovery is due to actual impairment of learning functions, time since stroke when initial testing is conducted, overall levels of illness, or whether recovery is a much slower process that is best observed over weeks and months rather than days. That is, the boundary conditions for observing recovery in the acute phase remain to be confirmed.

### **9.3.2 Cognitive Tasks**

The second aim of the research was to examine the possibility that cognitive tasks that varied the level of task difficulty might be more potent in detecting and/or facilitating improvement in performance. This aim was addressed in performance primarily on the Anagram task. Since performance is limited to a single test, it is also possible to explore the different ways in which learning has been assessed.

**9.3.2.1 Learning.** In earlier chapters the point was made that changes in z-scores across sessions were not necessarily a good indicator of learning because a stroke patient could still show improvement across sessions. If the improvement was not as great as that shown by the normative sample, however, the disparity between patient and sample could increase. With this caveat in mind, when one examines the changes in completion time on the Anagram task in terms of z-scores, MF and KS were in the clinical range on initial testing, and by Session 3 all four patients were in the clinical range. However, in their final session (Session 5 or Session 6) all patients with stroke were within normal range functioning.

Examination of test confidence intervals also indicated that there has been substantial improvement in performance. For KS, MF, and GL, where participants had experienced the single-event dual task condition for the second time, all had improved above baseline measures by the fourth session where only the primary task was tested. This improvement was maintained to the extent that in the final session where cognitive load was at its peak, performance was again outside the 95% confidence intervals. For these three participants there was significant improvement in performances even though there were increases in task difficulty across sessions. For these patients, improvement on this task was more pronounced than for the normative group. In the case of LL, there was no equivalent outcome because the increases in cognitive load tended to make more of an impact than was the case with the other patients.

The same outcome was apparent when scores were transformed into scaled scores. For KS, MF and GL, performance was within normal limits by Session 4 and was still within normal limits in their final session where cognitive load was at its maximum.

**9.3.2.2 Dual task effects.** Dual task effects were limited to performance on the Anagram task for all participants except for GL, who also provided data on the Snap task. The difficulty of the dual task was manipulated with a single-event detection task introduced in Session 3, and a two-event detection task introduced in Session 5. It was expected that the introduction of a secondary task could produce a number of effects on performance. There could be a slowing in the completion time on the primary task, an increase in errors on the primary task, and errors on the event detection task. In the normative sample, dual task effects were detectable only in one measure, the slowing in completion time on the primary measure when the two-event condition was introduced in Session 5.

Dual task effects were more pervasive in the stroke group. Like the normative sample, all patients with stroke showed slowing in completion time in Session 5, with LL also showing slowed completion time at Session 3 when the single-event detection time was introduced for the first time. Session 5 was also characterised by a spike in anagram errors compared to the previous and following sessions. Errors on the detection task were not problematic for either GL or MF. The same was true of LL except for the first session when the dual task was introduced. In this session it appeared that the instructions to detect events had not been processed and/or maintained. KS had major problems with this requirement; he consistently failed to detect the nominated events. In short, the performance of the stroke group was qualitatively the same as the control group in most aspects of performance, including the ability to monitor and process two distinct types of information at the same time.

To the extent that dual task methodologies reflect divided attention processes, the current results are consistent with a systematic review concerning the effectiveness attention training (Loetscher & Lincoln, 2013). That review suggested that the only compelling evidence for the effects of attention training were evident in divided attention tasks where dual task methodologies were utilised. Stablum et al. (2000) also used a dual task methodology in a working memory training study for patients with stroke in the chronic phase. They found consistent improvement across training sessions and considerable pre-test post-test scores on the training task. More importantly,

they found that this improvement was still maintained at 12 months, and that at post-test there was evidence of transfer to other cognitive tasks. The current results are entirely consistent with the training component of the task, but whether or not such gains are maintained over the long-term or show transfer effect are yet to be determined.

**9.3.2.3 Benchmark effects.** Anagram length is the final feature of this task that defines normal performance. In the normative sample, completion times for 7-letter anagrams were initially much faster than for 6-letter anagrams, and both were faster than 5-letter anagrams. With repeated testing, the length effect was attenuated with marked improvements in the speed of 5-letter solution time. Of the four patients with stroke who completed the Anagram task, all showed identical qualitative patterns to the normative sample. Thus, there were large effects of anagram length in one or more of the first three sessions, however the remaining sessions in each case were characterised by weakened length effects.

Performance on the Anagram task as measured by changes in absolute levels of performance, dual task effects, and the attenuation of anagram length, all indicated that the patients with stroke were performing this task without substantial impairment by the end of testing. This is in stark contrast to performance on the neuropsychological tests where only KS showed accelerated improved performance on impaired tasks. Even with such rapid improvement KS's performance was still within the clinical range on the final test session. Thus, the cognitive task passes the strong test of improvement that performance can return to within normal limits in the early days following stroke. This outcome suggests the possibility that cognitive tasks, where task difficulty is manipulated, are more potent at facilitating recovery than the simple repetition of tasks that are derived from neuropsychological test batteries. Consequently, the results provide some initial evidence that adaptive brain training exercises could facilitate recovery when practiced in the acute phase post-stroke.

## **9.4 Theoretical Implications**

### **9.4.1 Assessment**

Given that a number of neuropsychological tests were employed in the battery, it is logical that the current data could be used to address issues that are typically addressed in neuropsychological assessment. Up until this point in the thesis impairment has been determined at the task level and solely as a means of classifying performance. No reference was made to the neuropsychological domains that were affected by the stroke. This was done primarily to avoid issues outlined in Chapter 2 concerning the relationship between tasks and the constructs they measured (Kramer et al., 2014). However, one of the criteria used in the choice of tests to be employed in the battery was that they covered a number of domains that are typically used in neuropsychological assessment of stroke (Cumming et al., 2013; Jaillard et al., 2009; Nys et al., 2007; Planton et al., 2012; Tatemichi et al., 1994). It seems appropriate to address the issue of what domains have been impaired in the current data.

The problems of how tests and domains relate can be seen in reference to two of the tasks that were problematic for most participants namely the Stroop and semantic verbal fluency tasks. In Chapter 2 the results of a factor analysis conducted by Planton et al. (2012) are described (see Table 2.1). This large battery included the Stroop, phonemic fluency and semantic fluency among the tests they analysed. According to their results, the read and colour naming components of the Stroop task were measures of the mental speed of processing component of executive functions. The colour-naming component of the Stroop task was seen as measuring the inhibition component of executive functioning. From this perspective, the fact that difficulties were more pronounced on the read and colour naming component of the task would indicate speed of processing difficulties with fewer problems involving inhibition. The phonemic and semantic fluency tasks were interpreted as measures of the initiation component of executive functioning. The fact that there were different outcomes for phonemic and semantic fluency is problematic to the extent that deficits should be equivalent in the two tasks because they both measure the same construct.

An alternative method of accounting for the data involves the observation that in both the Stroop and verbal fluency tasks, the patients with stroke' deficits are most apparent on supposedly

easier components of the tasks rather than on harder tasks. Greater disparity on easier tasks is frequently found in the literature (Jaillard et al., 2009; Maidureira et al., 2001; Pustokhanova & Morozova, 2013; van Zandvoort et al., 2005).

In an attempt to understand these outcomes a more fine-grained analysis of the components of the Stroop and verbal fluency tasks was conducted where relative performance across conditions was explored rather than absolute performance. In Table 9.1 performance on the read and colour naming components can be compared by calculating the ratio of words read to colours named in the 45 second period. Likewise, read and colour naming components can be compared to the colour-word component by calculating the ratio of words produced in the read component relative to the colour-word scores, and the ratio of words produced in the colour naming component relative to the colour-word scores. The ratios are based on the average performance across the six sessions.

Table 9.1

*Ratios of Read/Colour-word Scores and Colour Naming/Colour-word Scores Averaged across Sessions*

Group/Patient	Read/Colour	Read/	Colour/
		Colour-Word	Colour-Word
Normative Sample	1.29	2.11	1.62
GL	1.22	2.01	1.64
MF	1.28	2.78	2.15
TB	1.07	1.09	1.03
KS	1.12	1.95	1.78
LL	1.32	2.89	2.40

The ratio derived from read and colour naming scores suggests that with the exception of TB, the remaining four patients with stroke do not differ markedly from the normative sample. Thus, although the patients with stroke are significantly slower on the reading and colour naming task than

the normative sample, there does not seem to be any impairment of the underlying cognitive processes. These results are thus consistent with the study by Planton et al.'s (2012) where performance on these tasks were deemed to be measures of speed of processing. However, when the ratios that involve colour-word naming are explored the patterns change. GL and KS show similar ratios to the normative sample, but MF and LL show much larger ratios reflecting increased difficulty on the colour-word naming task. Thus, one might conclude that again, although demonstrating speed of processing deficits, GL and KS show equivalent susceptibility to interference as the normative sample. On the other hand MF and LL experience greater levels of interference than the others. The fact that TB's scores remain constant can be attributable to the visual perception difficulties that he experienced.

In the case of GL and KS, the only difference in performance in comparison with the normative sample involves overall speed of processing. Their classification of severe impairment on the reading task but not on the colour-word task is due to the fact that slowed speed of processing has more of an impact on easy tasks than hard tasks. There is no evidence of impairment to the cognitive abilities that underpin this task. In the case of MF and LL, the same argument can be made as far as reading and colour naming goes. Their deficits in this area can again be attributed to speed of processing having more of an effect on the easier of the two tasks. However, for these two patients, the cognitive processes that protect against interference are impaired relative to the normative sample.

A similar set of analyses was conducted on the fluency tasks to explore impairment on the semantic fluency task among the patients with stroke. In these analyses scores on the first 15-second response period were compared to scores on the remaining 45-second period (Fernaes & Almkvist, 1998) on the assumption that different processes underpin performance at these intervals (Bittner & Crowe, 2006; Fernaeus & Almkvist, 1998; Fernaeus et al., 2008; Larsson et al., 2008). Table 9.2 presents the ratios of the number of later response items to first responses for both phonemic and semantic fluency tasks. A value above 1 indicates that more responses were made in the late

response condition than the early response condition, and a value of less than 1 indicates that more responses were made in the first 15-second period than in the subsequent 45-second period. For the normative sample, more words were generated in the later responses condition than the first responses condition and this was more pronounced with the semantic fluency task.

Table 9.2

*Ratios of Items Produced in the Later Response Period to First Response Period Averaged across Sessions*

	Phonemic Fluency	Semantic Fluency
Normative Sample	1.49	1.99
GL	1.03	1.05
MF	1.41	1.36
TB	.80	.52
KS	.94	.84
LL	.40	.44

With regards to the phonemic fluency task, only MF shows a ratio that approximates that of the normative sample, which is consistent with her classification as being not impaired in this task. In the case of the other four participants, the data indicate that they all experienced difficulty in producing items in the late response period. Impairment is not due to a speed of processing deficit;

there appears to be impairment in the underpinning cognitive abilities. In this case, slow and effortful retrieval from the lexicon (Fernaes & Almkvist, 1998).

In the case of the semantic fluency task, none of the patients with stroke have ratios that approximate the normative sample. Again, the problem appears to be the inability to generate items in the later response periods. What is more surprising is that for all patients with stroke the ratio of first to later responses is much the same for phonemic and semantic task. There appears to be a generalised problem in effortful retrieval from permanent knowledge bases (lexicon and semantic memory), but there is an additional issue as well in that whatever is changing performance in the normative sample between phonemic and semantic tasks is not influencing performance in the stroke group.

Table 9.3 presents the disparity in production between semantic and phonemic fluency tasks. In the normative sample, semantic fluency was the easier of the two tasks, with more items being produced on the semantic than phonemic task. In the case of normal performance the ratio of semantic items produced to phonemic items should be greater than 1. In the case where performance on the phonemic task was better than performance on the semantic task, the ratio would be less than 1. These ratios are presented in Table 9.3 in reference to the first responses period and the subsequent later responses period.

Table 9.3

*Ratios of Items Produced on the Semantic to Phonemic Fluency Tasks Averaged across Sessions*

Group/Patient	First Responses	Later Responses
Normative Sample	1.33	1.79
GL	1.22	1.26
MF	1.01	.93
TB	1.45	1.38
KS	.88	.71
LL	1.33	1.45

The normative sample shows the expected pattern of more items being produced on the semantic task than on the phonemic task (Fernaes & Almkvist, 1998), and this difference is more pronounced during the later responses. For the patients with stroke, GL, TB, and LL show the normal semantic advantage on the first responses. For MF, production is equivalent in the two tasks and for KS there is an obvious difficulty with semantic production. With regards to the later responses, all stroke participants do not produce the same change in disparity that characterises the normative sample, with the ratios again being quite similar to those observed in the first responses period.

As mentioned in Chapter 3, overall higher scores on the semantic task and the greater discrepancy between later and earlier responses on the semantic task have been attributed to differential strategy use. That is, in the late responses on the semantic test there is strong evidence for clustering items into sub-categories, which is not as evident in the first response period or on the phonemic fluency tasks (Raboutet et al., 2010; Troyer et al., 1998). One clear outcome suggested by the lack of change in ratios between semantic and phonemic production is that patients with stroke do not appear to take a strategic approach to the semantic fluency task. The fact that the ratios of later items to first items stays the same for both semantic and phonemic versions of the task is consistent with the non-use of strategies in the semantic version of the task.

The patients with stroke also differ from the normative sample on the rate of production in the first and later response period. With the possible exception of MF, all other patients experienced real problems in producing items during the later responses period. In this case the patients with stroke differ to other clinical groups where differences between groups are most pronounced during the first response period (TBI – Bittner & Crowe, 2006; Dementia - Ferneaus et al., 2008; Huntington's disease – Larsson et al., 2008).

While the primary aim of the thesis was not about determining what cognitive functions had been impaired, the above account of performance on the Stroop and verbal fluency tasks does

suggest that the measures used in the battery can be used to identify what cognitive processes have been impaired. Speed of processing deficits, difficulties in retrieval and the failure to use appropriate strategies have been identified as areas that were differentially problematic for the stroke participant in this study.

The above analyses also point to the utility of alternative ways of evaluating performance than simply relying on comparisons of total scores to normative behaviour. In the case of the Stroop task the ratio scores were useful in being able to distinguish between slowed speed of processing, which might simply be the result of being ill, and more fundamental cognitive impairments. Likewise, by dividing the verbal fluency tasks into early and late responses, it was possible to identify within-task difficulties and across task difficulties, and to attribute those difficulties to specific cognitive problems. Knowing what aspects of performance were impaired would provide useful assessment information in planning specific rather than rehabilitation programs (Ponsford et al., 2013; Taylor & Broomfield, 2013; Wilson, 2009).

#### **9.4.2 Learning**

The central issue in the research involved changes in performance across testing sessions. The first major issue addressed was whether or not consistent and sustained improvement could be detected. The data indicated that such behaviour could be observed and that recovery could be discriminated from standard practice effects when neuropsychological test were used. In short, the battery was sensitive enough to detect sustained improvement. However, the reality was that apart from KS, and to an extent MF, there was little evidence for sustained and consistent learning.

**9.4.2.1 Quality of data.** Standard advice regarding the administration of neuropsychological test batteries during the acute phase of stroke is that behaviour is so variable that accurate assessment of performance is unlikely to emerge (Lezak et al., 2004). Fluctuations in arousal, levels of physical illness, fatigue, depression, and anxiety have all been cited as reasons why there might be some doubt about the accuracy of scores on cognitive tests (Cumming, Brotmann, Darby, & Bernhardt, 2012; Lezak et al., 2004; van Zandvoort et al., 2005). The study by van Zandvoort et al. (2005)

confirmed the influence of these factors in that some tests could be completed by most of the patients with stroke in their studies, but many could not. However, they provide empirical evidence that reliable estimates of cognitive performance could be obtained during the acute phase with a large battery of tests, given that impairment could be detected and classified into unimpaired, mildly impaired, or severely impaired; that recovery could be detected on retesting between 12 and 24 months later; and testing during the acute phase was predictive of cognitive performance on the later testing. They also reported that arousal, fatigue and the other factors that were thought to have an unwanted impact on performance had no detectable effect on performance.

While the participants in the current study did report instances of tiredness and fatigue, it is not clear that this overly impacted on performance. If such factors did influence performance then relatively large variability across testing sessions would be expected, or there might be an across-task deficit on all tests when participants were having a bad day. On inspection of the variability in measures across sessions, it was apparent that the patients with stroke produced no more across-session variability than did many of the normative sample. Likewise, when looking at the composite scores, the only hint of an across-task change was a slight decrement in MF's performance on her final session, and LL's performance on Session 3. If these aberrant sessions are ignored, in both instances performance on the remaining tasks was quite stable. Thus, differential variability in test performance does not appear to provide an adequate explanation for the general lack of learning in the patients with stroke.

The current results reinforce the van Zandvoort et al. (2005) findings that reliable neuropsychological data can be obtained in the acute phase of stroke. Although the illness factors undoubtedly have an impact on performance, it is still possible to obtain data of a sufficient quality that learning effects can be evaluated. Accepted wisdom that testing during the acute phase is likely to not produce reliable data (Lezak et al., 2004) is questionable, at least as far as learning effects go.

The second issue involved the potency of neuropsychological tests and tasks derived from cognitive psychology training studies to facilitate change. In addressing this issue the cognitive rehabilitation and working memory training literature provided theoretical background to the development of the cognitive tasks.

**9.4.2.2 Facilitating biological recovery.** Robertson and Murre (1998) outlined five principles of direct cognitive rehabilitation training aimed at facilitating the biological recovery processes. Two of those processes had implication for the current study. The first was their principle of bottom-up specific stimulation. The idea was that tasks that were targeted at the specific damaged network should be frequently repeated to facilitate recovery in the damaged network. It is plausible that the repeated administration of neuropsychological tests could also provide similar bottom-up stimulation, which serves as a basis for accelerated learning. KS's performance on the impaired tasks is consistent with this account. However, it is also the case that this form of intervention was not overly efficient in that accelerated improvement was only evident in KS's performance. In the other patients there was little evidence of improvement. Consequently, the evidence supporting the effectiveness of bottom-up specific stimulation is not strong when it comes to cognitive impairments.

The second of Robertson and Murre's (1998) principles involved top-down specific stimulation. The expectation from this principle is that activating undamaged portions of a neural network should produce activation in other areas of the network including damaged areas. This principle was part of the theoretical underpinning of the Common Associates Test and the Immediate Serial Recall task. Unfortunately, because these tests did not meet the criteria for inclusion in the battery, this principle could not be assessed.

**9.4.2.3 Working memory training.** The working memory training literature served as the second theoretical basis for examining learning effects. This literature emphasises more proximate processes that are more directly linked with impairment, rather than focusing on neural determinants of performance (but see Klingberg, 2010, for an exception). The Vallat et al. (2005) research is a

good example of this approach where a specific deficit had been identified. The patient they tested had a specific working memory deficit, which they interpreted as deficits in the phonological loop component of Baddeley's (1986) working memory system. They devised a number of training tasks based on the phonological processes that purportedly underpin the phonological loop. Repeated practice on the component processes produced pre-test post-test improvements on a direct test of phonological loop functioning.

In the case of stroke where multiple impairments were expected, systematic reviews of the cognitive training literature (Loetscher & Lincoln, 2013) and individual studies (Stablum et al., 2000) indicated that tasks that employed a dual task methodology were effective in producing changes in behaviour. These methodologies have at their core the process of maintaining two sets of instructions in mind, and the ability to switch attention between tasks. As such, training on this fundamental process should theoretically produce widespread effects. Consequently, dual task components were built into two of the training tasks used in the current experiments and case studies. This manipulation was effective in that all participants showed dual task effects on the Anagram task, and all improved performance to the extent that they were within normal range functioning on their final test session. The current data confirm the potency of dual task methodologies in facilitating behaviour change.

The working memory training literature also indicated that tasks that required continuous updating of the contents of working memory were particularly effective in inducing improved performance on the task and in transfer effects (Jaeggi et al., 2008; Verhaeghen et al., 2004). The n-back task used in these studies formed the theoretical background for the use of the Snap task in the current research. GL was the only stroke patient who completed this task. On all the neuropsychological tests, both impaired and unimpaired, GL showed little evidence of learning. However, as is evident in Figure 8.1, GL's performance increased significantly across the first four sessions as indicated by the 95% confidence intervals. On the n-back task, therefore, GL exhibited

substantial and significant improvement whereas no such improvement was observed on the tasks that were not derived from the working memory training literature.

The working memory training literature was informative on what to train, and was equally informative on how to train. This training literature makes it clear that training has to be constantly challenging for the participant (Holmes et al., 2009; Klingberg et al., 2002, 2005). One way in which experimental and active control groups have been operationalised in a number of studies is to have the participants in each group perform the identical training tasks. For the active control group the task is maintained at the same level throughout the training period, but in the experimental group the level of difficulty is adapted from trial to trial so that the task is constantly difficult throughout the intervention period. While the current study did not vary performance in response to participants' scores, the progressive introduction of single-event and two-event conditions demonstrably increased task difficulty as per adaptive training requirements. Manipulating task complexity and task difficulty appears to have been a necessary and successful characteristic in facilitating improved performance on the Anagram tasks for the patients with stroke.

The fact that the outcomes involving the Snap and Anagram tasks are consistent with the working memory training literature provides the opportunity to speculate about interventions. Although the tasks were designed to evaluate differential behaviour change in neuropsychological and cognitive tasks, the repeated presentation of the two tasks under varying difficulty levels could be considered as a *de facto* training intervention. What has been demonstrated is that the patients with stroke have improved on the training task across test sessions. However, the strong test on the effectiveness of working memory training involves transfer effects (Chein & Morrison, 2010; Melby-Lervåg & Hulme, 2012; Morrison & Chein, 2011; Shipstead et al., 2012). Although there is strong debate about transfer effects, recent meta-analyses are converging on the recognition that near transfer effects can be found (Karchach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2012), but far transfer effects are less likely (Melby-Lervåg & Hulme, 2012; although see Karchach & Verhaeghen, 2014). In their meta-analysis of the working memory and executive function training studies,

Karbach and Verhaeghen (2014) suggested that working memory training could be broken up into three distinct classes, increasing working memory capacity, memory updating, and executive function. Among the executive functions they considered as a group were dual task performance, inhibition and interference control, task switching, and general forms of attention. From this perspective one might expect to see transfer from dual-task training to inhibition and interference control. In fact, some studies have shown that working memory training can result in transfer effects to the Stroop task, for young adults (Chein & Morrison, 2010), for children with ADHD (Klingberg et al., 2002, 2005), and older adults (Bherer et al., 2005; Karbach & Kray, 2009). In the case of the current case studies, improvements on the “training” tasks were not matched by improvements on the Stroop task for either GL or MF. The lack of transfer to the Stroop task reflected in the current data was also observed in the Westerberg et al. (2007) study, which suggests that while training benefits might be readily apparent, transfer effects may be more difficult to demonstrate in a stroke population.

## **9.5 Limitations**

### **9.5.1 Issues Associated with Levels of Impairment.**

The foundation assumption that determined expectations concerning learning and recovery was that a distinction between impaired and intact cognitive functions could be made. The conclusions that have been reached to this point have been based on the belief that the impaired/non-impaired decision was accurate. There are a number of factors that may impact on the decision making process.

A relatively trivial factor involves the selection of a cut-off on which to base the decision. For instance, using a cut-off z-score of -1.00 would lead to many more instances where performance was judged as impaired, just as a cut-off of -2.5 would produce a greater number of instances of non-impaired outcomes. In fact, in the literature reviewed in Chapter 2 there was a diversity of cut-offs used ranging from -1.67 to -2.00. However, the most frequently used criteria were a z-score of -1.96 (Tatemichi et al., 1994) or -2.00 (Jaillard et al., 2009; Hochstenbach et al., 1998; Planton et al.,

2012) and was consequently used in the current study. However, these cut-offs limit impairment to scores that are found in only 2.5% of the normal population. Some of the studies (Hurford et al., 2013; Nys et al., 2007; Sachdev et al., 2004a) used a cut-off on -1.67 in order to capture the bottom 5% of the population (also standard clinical cut-off). Adopting a criterion of -1.67 for assessing initial impairment would have made no substantial differences to the outcomes of the current research.

A second issue involves performance of the normative sample. All the comparisons have been made with reference to performance estimates drawn from a sample of 27 participants who varied across a relatively wide age range. This issue was addressed experimentally by including an age-matched control group, who did differ significantly from the larger group in terms of absolute levels of performance on some of the Stroop and semantic fluency measures. Thus for the patients with stroke, it is possible that initial assessment of impaired performance on these measures, where it occurred, may have been misclassified. More importantly, learning effects were equivalent in normative and control groups so inferences about learning effects are not compromised by comparisons to control or normative groups.

The appropriateness of the normative sample can also be addressed with respect to published norms. In standard assessment procedures, individual performance is typically compared to normative performance based on large samples where performance has been corrected for age and education levels. If the current normative sample is not representative of the larger samples, then patients with stroke might be misclassified as being impaired or not. To check this possibility, initial test performance of the normative sample was evaluated using published test norms for those tasks in which they were available. For each measure, the means of the normative group at initial testing were converted to scaled scores and compared to the published norms. A scaled score of 10 represents average performance in the published norms, and scaled scores of 11 converting to the 68<sup>th</sup> percentile. For digit span the normative group had scaled scores of 10 and 11 for forward and backward recall respectively (WAIS-IV, Wechsler, 2006), suggesting that the normative sample was

representative of the larger population. The same was true of the Rey Tangled Lines test where the scaled score was 10 (Senior et al., 1998) and the colour-word and read scores of the Stroop test with scores of 10 and 11 respectively (Ivnik et al., 1996). However, the normative sample produced better performance than the normative population on the colour naming component of the Stroop test with a score of 13 (Ivnik et al., 1996), the total scores on phonemic fluency task with a score of 13 (Tombaugh et al., 1999), and for total scores on the animal naming component of the semantic fluency task with a scaled score of 18 (Tombaugh et al., 1999). In the case of the verbal fluency measures in particular, the normative sample does differ quite markedly from the larger samples employed in published norms.

Given the differences between sample and published norms, patients with stroke who were identified as impaired based on the normative sample may well be within normal range functioning when the published norms are used to determine impairment. To examine this possibility, initial impairment of the patients with stroke using published norms was examined. Table 9.4 compares the classification of unimpaired performance based on z-scores using the normative sample and scaled scores based on published norms for the tasks under consideration.

Table 9.4

*Classification of Unimpaired Performance at Initial Test as a Function of Cognitive Test and Comparison Group*

Measure	Z-scores	SS(Published)
Digit Forward	GL,LL,MF,KS,TB	GF,LL, MF,TB,KS,
Digit Backwards	GL,LL,MF,KS,TB	GL,LL,MF,TB,KS
Stroop - Read		GL, MF,LL
Stroop - Colour		GL,MF
Stroop - ColWord	GL,LL, TB	GL, MF, TB,LL
FAS-Total	GL,MF,KS,LL	GL,MF,KS,LL
Animal- Total	MF	MF
Rey Tangled Lines	MF,KS	MF

As is evident in the Table, the systematic differences in classification are restricted to the Stroop test where GL, MF and LL are in the unimpaired range according to the published norms. On

the other studies, particularly the verbal fluency tasks where the discrepancies between sample and population are most pronounced, the classification is identical. Moreover, while these discrepancies in impairment classification on the Stroop task are interesting, they have no bearing on the primary outcomes of the study. Neither GL, MF nor LL showed any learning on any of the Stroop measures.

### **9.5.2 Issues Associated with Learning.**

In the absence of information concerning changes in behaviour in the acute phase of stroke, the current research can be viewed as a proof of concept exercise that explores the extent to which reliable behaviour change can be measured, standard practice effects could be discriminated from natural recovery, and whether or not different types of tasks were more potent in facilitating improvements in performance. While answers to the learning aspects have been provided, no attempt has been made to either explore why the effects emerged or not, or what the boundary conditions were for observing the effects.

An obvious question to ask was why KS showed accelerated recovery in the impaired tasks and the other patients did not. KS differed from the other patients in terms of his age and when testing began. A large limitation of the current research is that these issues cannot be addressed with the current data. For instance there are the obvious issues like the severity of the stroke, the type of stroke, and the location of the infarct. These issues (and others) have not been addressed in the current study, primarily because it was deemed important to establish that behaviour could improve before addressing issues of why performance did or did not change.

The current outcomes indicate that these issues should be explored in future research. On the basis of the current outcomes there are several factors that might place boundary conditions on observing improvements in behaviour.

One obvious reason for not observing recovery is the time frame over which testing was conducted. For most patients with stroke, testing was completed within a 2-week period from the time they were first admitted to the rehabilitation ward to the time they were discharged. While KS showed that recovery can be observed under these conditions, it may be the case that repeated testing

needs to be distributed over a wider time frame for recovery to be observed. More importantly, it may be the case that the best effects are observed when initial testing is delayed for a period. There is currently some debate as to how early “early intervention” should be. It has been argued that if therapy is initiated too early, it can be ineffective at best, but can also have detrimental effects (Bernhardt, Indredavik, & Langhorne, 2012).

Another limitation to understanding change in the current patients with stroke involves the lack of good pre-morbid estimates of functioning. While pre-morbid estimates of cognitive functioning were calculated on the basis of sex, age and education the confidence bands associated with these estimates were very large. One of surprising findings was the lack of improvement on the tests that were classified as unimpaired. While performance on an initial test was in the normative range, it is still possible that the stroke could have produced a significant deterioration on the tests if the patients with stroke performed in the above average range on these tests prior to the stroke. While it is unlikely that this is a widespread occurrence in the current data, in the absence of better information about pre-morbid functioning, certainty about intact cognitive domains cannot be assured. While better estimates of pre-morbid functioning could have been included in the test battery, this would have increased test-time, the size of the battery, and the cognitive effort on the part of the stroke patient. For practical reasons, preference was given to the other tests in the battery that were directly relevant to the recovery process.

The lack of learning might also be related to the number or type of comorbidity experienced by each of the patients with stroke. As an example, depression and anxiety are comorbidities of stroke in 20% to 25% of cases (Hachinski, 1999; Schramke, Stowe, Ratcliff, Goldstein, & Condray, 1998; van Zandvoort et al., 2005) and can have a negative influence of cognitive test performance (Barker-Collo, Feigin, Parag, Lawes, & Senior, 2010). One possibility for the lack of learning could be that learning is one of the cognitive functions that is attenuated with high levels of depression (or other comorbidities). While this is a plausible explanation, there are instances in the literature where participants with clinical depression show normal learning performance even though there are

differences in absolute levels of performance compared to non-depressed controls (Hammar & Ardal, 2012; Weingartner et al., 1993). While it is highly likely that multimorbidities are present in the patients with stroke (Gallacher et al., 2014), it is not clear that any combination will produce a decrement in performance, and it is even less clear as to whether or not learning is affected by the combination of comorbidities. In addition, to rely on a comorbidity argument for the lack of learning one would need to assume that the effect size of a pre-existing condition swamps the effects due to a recently experienced stroke.

### **9.5.2 Issues Associated with Sample Size**

A case study approach, as opposed to a large-n study, was the preferred approach in evaluating the research aims. There were a number of reasons for adopting such an approach. The high likelihood of heterogeneity in the number and type of cognitive impairment (Tatemichi et al., 1994; Hurford et al., 2013), together with equivalent heterogeneity in multimorbidities (Caughey et al., 2010; Gallacher et al., 2014) meant that controlling for all probable confounds was not realistically achievable (Stuart & Rubin, 2007). Controlling for the possible effects of unwanted variables is an essential aspect of large-n designs aimed at evaluating experimental group/control group differences. Time limits in collecting data, recruitment and drop-out issues, along with hospital priorities in terms of processes and procedures were practical reasons for adopting the low-n approach. The appropriateness of a case study methodology for informing future rehabilitation needs of the patient was also a motivating factor. Case studies typically provide more detailed descriptions of performance at the individual level, which can form the basis of rehabilitation programs (Caramazza, 1986; Wilson, 2009), whereas the large-n studies describe average performance where it is possible that no individual participant has the precise characteristics of the “average” participant.

One large disadvantage of the case study approach involves the extent to which the outcomes of a small number of participants are representative of the wider population. With a case study approach it is not possible to make strong inferences concerning what might happen with other

participants. Even with a multiple case study approach as adopted here, the question of how many case studies are required to capture the essential variation among members of the population as a whole is an unresolvable problem without conducting large-n studies.

In spite of the recognised generalisation problems associated with the current methodology, there are a number of outcomes from the current case studies that can inform future large-n studies. One logical assumption made concerning possible outcomes was that four logical profiles were possible: improved performance would be seen in both initially impaired and unimpaired domains; improved performance would be seen on the unimpaired domains and no improvement would be observed on the impaired domains; improved performance would be observed on impaired domains and no improvement would be found on intact domains; and there would be no improvement on any of the domains.

While these profiles are logically possible, they are based on the assumption that all tasks deemed to be unimpaired at baseline would change in the same way with repeated exposure, and all those that were initially impaired would change in the same way across testing sessions. However, a possible outcome was that some of the unimpaired tasks would improve and some would not, and some of the impaired tasks would improve and some not. That is, improvement or lack thereof would not be coupled to initial levels of impairment. The fact that this did not occur and that three distinct profiles among the neuropsychological tests were discernible has implications for future studies. If the current results can be replicated, large-n studies can largely ignore the issues associated with heterogeneity of impairments and concentrate on determining the prevalence of the four possible profiles without having to attend to specific tasks or specific domains.

The fact that profiles were identifiable both in patients with no known comorbidities and those that do had identified co-morbidities also has implications for further studies. Again, the suggestion would be that multimorbidities might be associated with a particular type of profile, but it is still the case that the emphasis should be on learning profiles among impaired and unimpaired domains, not the tasks themselves.

Whether or not similar profiles exist on the cognitive tasks cannot adequately be addressed from the current case studies given that cognitive task performance was effectively limited to performance on a single task. Before large-n studies could be completed, learning would need to be evaluated using a number of cognitive tasks to demonstrate that improvement was not a factor of specific tasks, but was sensitive to the degree of initial impairment. The current finding that all patients with stroke who were able to do the cognitive task showed improvement, provides some initial data indicating that the prevalence rates for the four possible profiles may be different for neuropsychological and cognitive tasks.

### **9.6 Future Studies**

In evaluating the outcomes of the studies, some suggestions have been made concerning future research in the area, particularly with respect to the boundary conditions associated with learning. Testing of current patients with stroke took place in a 2- to 3-week window. While this period was sufficient to show recovery effects in at least one patient, it may be too brief a period to track natural recovery in the majority of patients.

There is obvious work to be done in test selection and test construction. Because the battery serves a function of assessing initial levels of cognitive function, the battery must be large enough to cover the cognitive domains that are most likely to be impaired by stroke (Hurford et al., 2013). Whether or not the existing battery should continue to be used or adapted is worth considering. The results show that some of the tests are more effective at showing change than others. In the case of the Stroop and verbal fluency tasks, improvement was apparent in the normative group in the very early stages of testing (Session 2). In the case of backward digit span, detectable improvement only emerged on the third testing session, and for digit forward and Rey Tangled Lines, improvement only emerged in the later sessions. Thus, there is some question about the utility of forward span and Rey Tangled Lines for this purpose. In the case of forward span the reliability of the test is the least impressive of all the tasks, and this may be the root cause of the late detection of improvement. The Rey Tangled Lines Test did not suffer from reliability issues so the late emergence of learning

effects appears to be caused by other unknown factors. The Saccadic Tracking Test (Reischies & Berghofer, 1995) might be one possible replacement in that, like the Rey Tangled Lines test, it is a brief task that requires visual processing of a specified trail and does not involve any motor component. If greater learning was present in this task relative to the Rey Tangled Lines test, then it would be worth using it instead of the Rey Tangled Lines test.

The aim of using spreading activation as a means of facilitating performance was not realised in the current battery. It might still be possible to develop alternative training tasks in which this aspect of normal cognition could be exploited. Thus, priming in lexical decision or naming tasks could be utilised, or adaptations of the DRM paradigm (Roediger & McDermott, 1995) could be exploited. Alternatively, given the success of the dual-task paradigm in the current studies, it might be preferable to develop some other simple dual-task measures, or to adopt other working memory type tasks, like the n-back task, where task difficulty can be consistently monitored and manipulated. In sum, the current data suggests that the Stroop task, the verbal fluency measures, the anagram task and the snap task could form the basis of future studies. Given that the cognitive task(s) produced the better outcomes, additional dual-task measures could and should be developed. One of the clear outcomes of the current research was the difficulties the patients with stroke experienced with the semantic fluency task. The development of a targeted training program around this area of performance appears to be particularly pressing. Again, across-task priming could be implemented. Alternatively, short periods (15-seconds) using sub-categories as explicit cues (e.g., farm animals, wild animals, Australian animals, etc.) could provide direct practice on the tasks.

In Chapter 2 when the aims and objectives were outlined, one of the points made was that before issues involving demographic or neurological factors could be considered, the utility of the battery had to be established. It was argued that utility could only be confirmed if four conditions were met. Those were: a) stable performance could be obtained with repeated testing; b) that consistently impaired and unimpaired performance could be discriminated; c) consistent improvement could be observed across test sessions; and d) differences between neuropsychological

and cognitive tests would be observed. Given that these four conditions can be met in the acute phase, it now becomes potentially worthwhile to consider some of these factors. One obvious question to explore in future studies is the link between recovery (or its absence) with the size and location of the infarct. Moreover, if training regimes were implemented based upon the current findings, then it would also make sense to explore location effects on training outcomes as well as on recovery.

### **9.7 Conclusions**

The aftermath of a stroke can result in impairments in motor, sensory, language and other cognitive processes. However, in the days following stroke, biological repair mechanisms commence with the result that recovery in the impaired behaviours can be observed. In the motor and language areas, therapeutic interventions that have been implemented in the early days and weeks following a stroke have been shown to facilitate recovery and produce long-lasting functional benefits. Examination of the literature concerning recovery of cognitive deficits and corresponding interventions show two areas where there is a substantial gap in the literature. Firstly, there are no published studies that have tracked the progress of cognitive recovery during the acute phase of stroke. Secondly, there are no published studies that explore interventions aimed at facilitating recovery during this same period. Consequently, the aim of the current research was to address these issues by repeatedly administering a battery of cognitive tests across six sessions, within a 2- to 3-week period, to people who had recently suffered a stroke.

The test battery involved standardised neuropsychological tests that assessed a variety of cognitive domains, and served the role of assessing cognitive recovery. The battery also contained cognitive tests in which task difficulty was manipulated by the introduction of dual task conditions. These tasks explored the possibility, derived from the working memory training literature, that cognitive tests might be more potent in detecting or facilitating improved performance. The battery was repeatedly administered to a group of participants who had not suffered a stroke but who were

in the age range that covered 85% of stroke cases in Australia. The battery was administered to this group in order to establish or confirm the psychometric properties of the tests in the battery. They also provided normative information concerning absolute levels of performance on the various tasks, and more importantly how performance changed with repeated exposure to the test battery and with the introduction of dual task components. The outcomes of this component confirmed that, for most tasks and most measures, performance did improve with repeated exposure, although such improvement emerged in later testing sessions for some tasks and in earlier testing sessions for other tasks. This normative sample also confirmed that performance on the intervention tasks was sensitive to the implementation of dual task conditions.

Expectations about the learning outcomes associated with having a stroke depended on first determining whether or not performance on each task was within normative limits. When performance was not impaired, the expectation was that the patients with stroke would show the same learning effects that were observed in the normative sample. When indications were that performance on the task had been impaired by the stroke, the expectation was that natural recovery would be indicated by improved scores with repeated testing. In the case where such improvement was not detected, the assumption was that recovery was either not taking place or taking place at a much slower rate, which was not detectable under the conditions of the testing regime.

There were a number of outcomes of the study, some expected and others not. In the first instance, the results indicated that it was possible to obtain valid and reliable data (van Zandvoort et al., 2005). As is evident in much of the stroke literature (Tatemichi et al., 1994), the patients showed evidence of multiple impairments that covered a number of different cognitive domains. Just as importantly, all patients showed that some aspects of cognition had been spared, in that performance on the relevant tasks was within normal limits. In regards to tracking recovery via the neuropsychological tests, the changes in performance exhibited by KS indicated that recovery of impaired function could be observed with the adopted testing regime. However, KS was the only stroke patient who demonstrated such recovery. There was very little evidence for cognitive

recovery during the 2- to 3-week testing period among the other patients with stroke. Furthermore, three of these patients did not show consistent and sustained improvement on tasks where initial performance had indicated impaired functioning, or on tasks where functioning was judged to be unimpaired.

The more positive outcome of the research involved changes as assessed by the cognitive test. All patients with stroke were in normal range functioning on the primary task measure by the end of testing, all showed evidence of dual task effects, and all showed the same changes in benchmark effects that were observed in the normative sample. Although the cognitive component of the battery was restricted to a single task, the results do provide some support for the notion that cognitive tasks are more potent in assessing or facilitating recovery during the acute phase of stroke. A tentative implication of this finding is that cognitive interventions in the acute phase of stroke may facilitate the recovery process, just as early interventions in the motor and language domain have been shown to facilitate recovery (Wolf et al., 2008; Meinzer et al., 2007).

The research process has resulted in a number of contributions. Firstly, the two gaps in the literature that were identified in the earlier chapters have been addressed. The outcomes of the current research indicate that it is possible to track the recovery process during the acute phase of stroke with repeated administration of neuropsychological test batteries, and that cognitive tests are potentially more effective in this same period. The research also makes a contribution in terms of identifying the manipulation of task difficulty as a key component in the development of future cognitive interventions. The research has introduced new measures on the verbal fluency tasks that have proven useful in diagnosing the locus of impaired processing. The introduction and use of ratio measures between task components in the Stroop and verbal fluency tasks were also useful in assessing changes in performance and in diagnosing the cause of impairment. The utility of these measures is not restricted to the stroke populations; they can just as readily be applied to any areas where cognitive functioning is suspected of being impaired. The changes in performance on the Anagram task also demonstrated the utility of building benchmark phenomenon into the cognitive

tasks used in the battery as a means of establishing convergent data on the learning process. Finally, the research has contributed to the way in which learning can be assessed at the individual level using multiple measures of learning that provide complementary and convergent information on the learning process.

In the early days following a stroke, biological repair processes are initiated that produce some degree of behavioural recovery. The research reported here leads to the cautious conclusions that cognitive recovery can be measured during this acute phase, and that future cognitive interventions based on the principles that govern working memory training programs may be effective in facilitating recovery. As this is a first attempt at exploring these issues, further research is required to replicate the current findings and to explore the parameters that influence change. If the research can be replicated, long-term outcomes of early intervention need to be assessed both in terms of the maintenance of cognitive performance, but to broader functional behaviours that determine quality of life. There are enough positive outcomes in the current research to indicate that this further research should be conducted.

### References

- Albert, S. J. & Kesselring, J. (2012). Neurorehabilitation in Stroke. *Journal of Neurology*, 259, 817-832. doi: 10.1007/s00415-011-6247-y
- Amunts, K., Schlaug, G., Jancke, L., Steinmetz, H., Schleicher, A., Dabringhaus, A., & Zillus, K. (1998). Motor cortex and hand motor skills: Structural compliance in the human brain. *Human Brain Mapping*, 5, 206-215. doi: 10.1002/(SICI)1097-0193(1997)5:3<206::AID-HBM5>3.0.CO;2-7
- Anstey, K. J. & Low, L-F. (2004). Normal cognitive changes in aging. *Australian Family Physician*, 33, 783-787.
- Appelros, P. (2005). Characteristics of mini-mental state examination 1 year after stroke. *Acta Neurologica Scandinavica*, 112, 88-92. doi: 10.1111/j.1600-0404.2005.00441.

- Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gomez, C. (2000). Age-related cognitive decline during normal aging: The complex effect of education. *Archives of Clinical Neuropsychology, 15*, 495-513. doi: 10.1016/S0887-6177(99)00040-2
- Ardila, A., Ostrosky-Solis, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of semantic verbal fluency (ANIMALS). *International Journal of Psychology, 41* (5), 324-332. doi: 10.1080/00207590500345542
- Ashcraft, M. J. (2005). *Cognition* (4<sup>th</sup> edition). USA: Prentice Hall.
- Australian Institute of Health and Welfare. (2013). Stroke and its management in Australia: an update. Cardiovascular disease series no. 37. Cat. no. CVD 61. Canberra: AIHW
- Axelrod, B. N., Torner, R., Fisher, T., & Aharon-Peretz, J. (2001). Preliminary analyses of Hebrew verbal fluency measures. *Applied Neuropsychology, 8*, 248-250. doi: 10.1207/09084280152829101
- Azuma, T., Cruz, R. F., Bayles, K. A., Tomoeda, C. K., & Montgomery, E. B. (2003). A longitudinal study of neuropsychological changes in individuals with Parkinson's disease. *International Journal of Geriatric Psychiatry, 18*, 1043-1049. doi: 10.1002/gps.1015
- Baddeley, A. D. (1986). *Working Memory*. Oxford University Press: Oxford.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 11*, 417-423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation, 8*, 47-89.
- Ballard, C., Rowan, E., Stephens, S., Kalaria, R., & Kenny, R. A. (2003). Prospective follow-up study between 3 and 15 months after stroke: Improvements and decline in cognitive function among dementia-free stroke survivors >75 years of age. *Stroke, 34*, 2440-2444. doi: 10.1161/01.STR.0000089923.29724.CE

- Barker-Collo, S., Feigin, V. L., Parag, V., Lawes, C. M., & Senior, H. (2010). Auckland Stroke Outcomes study: Part 2: Cognition and functional outcomes 5 years poststroke. *Neurology*, *75*(18), 1608-1616. doi: 10.1212/WNL.0b013e3181fb44c8
- Bechtoldt, H., Benton, A., & Fogel, M. (1962). An application of factor analysis in neuropsychology. *Psychological Record*, *12*, 147-156.
- Bernhardt J, Indredavik B, Langhorne P. (2012). When should rehabilitation begin after stroke? *International Journal of Stroke*, *8*, 5-7. doi: 10.1111/ij.s.12020
- Berthier, M. L., Green, C., Higuera, C., Fernandez, I., Hinojosa, J., & Martin, M. C. (2006). A randomized, placebo-controlled study of donepezil in poststroke aphasia. *Neurology*, *67*, 1687-1689. doi: 10.1212/01.wnl.0000242626.69666.e2
- Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2005). Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control? *Psychology and Aging Special Issue: Emotion-Cognition Interactions and the Aging Mind*, *20*, 695– 709. doi: 10.1037/0882-7974.20.4.695
- Biernaskie, J., Chernenko, G., & Corbett, D. (2004). Efficacy of rehabilitative experience declines with time after focal Ischaemic brain injury. *The Journal of Neuroscience*, *24*, 1245-1254.
- Binkofski, F., Seitz, R., Arnold, S., Classen, J., Benecke, R., & Freund, H. J. (1996). Thalamic metabolism and corticospinal tract integrity determine motor recovery in stroke. *Annals of Neurology*, *39*, 460–470. doi: 10.1002/ana.410390408
- Bittner, R. M. & Crowe, S. F. (2006). The relationship between naming difficulty and FAS performance following traumatic brain injury. *Brain Injury*, *20*, 971-980. doi: 10.1080/02699050600909763
- Blake, H., McKinney, M., Treece, K., Lee, E., & Lincoln, N. (2002). An evaluation of screening measures for cognitive impairment after stroke. *Age and Ageing*, *31*, 451-456. doi: 10.1093/ageing/31.6.451

- Bohnen, N. I., Kotagal, V., Müller, M. M., Koeppe, R. A., Scott, P. H., Albin, R. L., & ... Petrou, M. (2014). Diabetes mellitus is independently associated with more severe cognitive impairment in Parkinson disease. *Parkinsonism & Related Disorders*, *20*, 1394-1398. doi: 10.1016/j.parkreldis.2014.10.008
- Bott, N. T., Johnson, E. T., Schuff, N., Galifianakis, N., Subas, T., Pollock, J., & ... Possin, K. L. (2014). Sensitive measures of executive dysfunction in non-demented Parkinson's disease. *Parkinsonism & Related Disorders*, *20*, 1430-1433. doi: 10.1016/j.parkreldis.2014.10.007
- Bour, A., Rasquin, S., Boreas, A., Limburg, M., & Verhey, F. (2010). How predictive is the MMSE for cognitive performance after stroke? *Journal of Neurology*, *57*, 630-637. doi: 10.1007/s00415-009-5387-9
- Bowden, E. M. & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behaviour Research Methods, Instruments and Computers*, *35*, 634-639. doi: 10.3758/BF03195543
- Bower, G. H. & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science*, *20*, 119-120.
- Brayne, C., Nickson, J., McCracken, C., Gill, C., & Johnson, A. L. (1998). Cognitive function and dementia in six areas of England and Wales: The distribution of MMSE and prevalence of GMS organicity level in the MRC CFA study. *Psychological Medicine*, *28*, 319-335. doi: 10.1017/S0033291797006272
- Brion, J-P., Demeurisse, G., & Capon, A. (1989). Evidence of cortical reorganization in hemiparetic patients. *Stroke*, *20*, 1079-1084.
- Buhner, M., König, C. J., Prick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, *19*, 253-275. doi: 10.1207/s15327043hup1903\_4

- Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-Analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*, 543-570. doi: 10.1080/13854046.2012.680913
- Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test–retest correlations. *The Clinical Neuropsychologist*, *27*, 1077-1105. doi: 10.1080/13854046.2013.809795
- Calero, M. D. & Navarro, E. (2004). Relationship between plasticity, mild cognitive impairment and cognitive decline. *Archives of Clinical Neuropsychology*, *19*, 653–660. doi: 10.1016/j.acn.2003.08.008
- Calero, M. D. & Navarro, E. (2007). Cognitive plasticity as a modulating variable on the effects of memory training in elderly persons. *Archives of Clinical Neuropsychology*, *22*, 63–72. doi: 10.1016/j.acn.2006.06.020
- Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition*, *5*, 41–66. doi: 10.1016/0278-2626(86)90061-8
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386-404. doi: 10.1016/0022-0965(82)90054-6
- Cassidy, T. P., Lewis, S., & Gray, C. S. (1998). Recovery from visuospatial neglect in stroke patients. *Journal of Neurology Neurosurgery and Psychiatry*, *64*, 555-557. doi: 10.1136/jnnp.64.4.555.
- Caughey, G. E., Ramsay, E. N., Vitry, A. I., Gilbert, A. L., Luszcz, M. A., Ryan, P.,..Roughhead, E. E. (2010). Comorbid chronic diseases, discordant impact on mortality in older people: a 14-year longitudinal population study. *Journal of Epidemiology and Community Health*, *64*, 1036-1042. doi: 10.1136/jech.2009.088260.

- Chein, J. & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin and Review*, *17*, 193-199. doi: 10.3758/pbr.17.2.193
- Chen, R., Chen, E., Chan, C., Lam, L., & Lie-Mak, F. (2000). Verbal fluency in schizophrenia: Reduction in semantic store. *Australian and New Zealand Journal of Psychiatry*, *34*, 43-48. doi: 10.1046/j.1440-1614.2000.00647.x
- Chouinard, P.A., Leonard, G., & Paus, T. (2006). Changes in effective connectivity of the primary motor cortex in stroke patients after rehabilitative therapy. *Experimental Neurology*, *201*, 375-387. doi: 10.1016/j.expneurol.2006.04.037
- Chung, C., Pollock, A., Campbell, T., Durward, B., & Hagen, S. (2013). Cognitive rehabilitation for executive dysfunction in adults with stroke or other adult Nonprogressive acquired brain damage. *Stroke*, *44*, e77-e78. doi: 10.1161/STROKEAHA.113.002049
- Cicerone, K. D., Harley, J. P., Laatsch, L., Morse, P. A., Catanese, J., Dahlberg, C., ... & Giacino, J. T. (2005). Evidence-based cognitive rehabilitation: Updated review of the literature from 1998 through 2002. *Archives of Physical Medicine and Rehabilitation*, *86*, 1681-1692. doi: 10.1016/j.apmr.2005.03.024
- Cicerone, K. D., Bergquist, T., Azulay, J., Cantor, J., Ashman, T. Langenbahn, D. M., ... & Harley, J. P.. (2011). Evidence-based cognitive rehabilitation: Updated review of the literature from 2003 through 2008. *Archives of Physical Medicine and Rehabilitation*, *92*, 519-530. DOI: 10.1016/j.apmr.2010.11.015
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428. doi: 10.1037/0033-295X.82.6.407
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *8*(2), 240-247. doi: 10.1016/S0022-5371(69)80069-1

- Cooper, D., Epker, M., Lacritz, L., Weine, M., Rosenberg, R., Hongig, L., & Callum, C. (2001). Effects of practice on category fluency in Alzheimer's Disease. *The Clinical Neuropsychologist, 15*, 125-128. doi: 10.1076/clin.15.1.125.1914
- Cramer, S. C. (2008). Repairing the human brain after stroke: 1. Mechanisms of spontaneous recovery. *Annals of Neurology, 63*, 272-287. doi: 10.1002/ana.21393
- Cramer, S. C. (2011). An overview of therapies to promote repair of the brain after stroke. *Head & Neck, 33*, S5-S7. doi: 10.1002/hed
- Cramer, S.C., Shah, R., Juraneck, J, Crafton, K. R., & Le, V. (2006). Activity in the peri-infarct rim in relation to recovery from stroke. *Stroke, 37*, 111–115. doi: 10.1161/01.STR.0000195135.70379.1f
- Crosson, B., Moore, A. B., Gopinath, K., White, K. D., Wierenga, C. E., Gaiefsky M. E.....Gonzalez Rothi, L. J. (2005). Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. *Journal of Cognitive Neuroscience, 17*, 392-406. doi: 10.1162/0898929053279487
- Crowe, S. (1992). Dissociation of two frontal lobe syndromes by a test of verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 14*, 327-339.
- Crutch, S. J. & Warrington, E. K. (2008). Contrasting patterns of comprehension for superordinate, basic-level, and subordinate names in semantic dementia and aphasic stroke patients. *Cognitive Neuropsychology, 25*, 582-600.
- Cumming, T. B., Brodtmann, A., Darby, D., & Bernhardt, J. (2012). Cutting a long story short: Reaction times in acute stroke are associated with longer term cognitive outcomes. *Journal of the Neurological Sciences, 322*(1-2), 102-106. doi: 10.1016/j.jns.2012.07.004
- Cumming, T. B., Marshall, R. S., & Lazar, R. M. (2013). Stroke, cognitive deficits, and rehabilitation: Still an incomplete picture. *International Journal of Stroke, 8*, 38-45. doi: 10.1111/j.1747-4949.2012.00972.x

- Cumming, T. B., Thrift, A. C., Collier, J. M., Churilov, L., Dewey, H. M., Donnan, G. A., & Bernhardt, J. (2011). Very early mobilization after stroke fast-tracks return to walking: Further results from the Phase II AVERT Randomized Controlled Trial. *Stroke*, *42*, 153-158. doi:10.1161/STROKEAHA.110.594598
- Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510-1512. doi: 10.1126/science.1155466
- Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*(4), 720-730. doi: 10.1037/a0014296
- Dancause, N. J., Barbay, S., Frost, S. B., Plautz, E. J., Chen, D., Zoubina, E. B., Stowe, A. M., & Nudo, R.J. (2005). Extensive cortical rewiring after brain injury. *Journal of Neuroscience*, *25*, 10167-10179. doi: 10.1523/JNEUROSCI.3256-05.2005
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6
- Darby, D., Maruff, P., Collie, A., & McStephen, M. (2002). Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology*, *59*, 1042-1046.
- Desmond, D. W., Moroney, J. T., Sano, M., & Stern, Y. (1996). Recovery of cognitive function after stroke. *Stroke*, *27*, 1798-1803. doi: 10.1161/01.STR.27.10.1798
- Dick, J.P., Guiloff, R. J., Stewart, A., Blackstock, J., Bielawska, C., Paul, E. A., & Marsden, C. D. (1984). Mini-mental state examination in neurological patients. *Journal of Neurology, Neurosurgery and Psychiatry*, *47*, 496-499. doi: 10.1136/jnnp.47.5.496
- Dimitrov, M., Granetz, J., Peterson, M., Hollnagel, C., Alexander, G., & Grafman, J. (1999). Associative learning impairments in patients with frontal lobe damage. *Brain and Cognition*, *41*, 213-230. doi:10.1006/brcg.1999.1121

- Donovan, K., Siegert, R., & McDowall, J. (1999). Clustering and switching in verbal fluency in Parkinson's disease. *New Zealand Journal of Psychology, 28*, 61-66.
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdhan, U., & May, A. (2004). Changes in grey matter induced by training. *Nature, 427*, 311-312. doi: 10.1038/427311a
- Duan, X., Long, Z., Chan, H., Liang, D., Qui, L., Huang, X.....Gong, Q. (2014). Functional organization of intrinsic connectivity networks in Chinese-chess experts. *Brain Research, 1558*, 33-43. doi: 10.1016/j.brainres.2014.02.033
- Duncan, P. W., Goldstein, L. B., Horner, R. D., Landsman, P. B., Samsa, G. P., & Matchar, D. B. (1994). Similar motor recovery of upper and lower extremities after stroke. *Stroke, 25*, 1181-1188.
- Duncan, P. W., Goldstein, L. B., Matchar, D., Divine, G. W., & Feussner, J. (1992). Measurement of motor recovery after stroke: Outcome assessment and sample size requirements. *Stroke, 23*, 1084-1089.
- Duff, K., Callister, C., Dennet, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neurologist, 26*, 1117-1127. doi: 10.1080/13854046.2012.722685
- Duff, K., Beglinger, L. J., Moser, D., Schultz, S., & Paulsen, J. S. (2010). Practice effects and outcome of cognitive training: Preliminary evidence from a memory training course. *American Journal of Geriatric Psychiatry, 18*, 91.
- Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S.,..... & McCaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnesic mild cognitive impairment. *American Journal of Geriatric Psychiatry, 19*, 932-939. doi: 10.1097/JGP.0b013e318209dd3a
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309-331. doi: 10.1037/0096-3445.128.3.309

- Faustman, W. O., Moses, J. A., & Csemansky, J. G. (1990). Limitations of the Mini-Mental State Examination in predicting neuropsychological functioning in a psychiatric sample. *Acta Neurologica Scandinavica*, *81*, 126-131. doi: 10.1111/j.1600-0447.1990.tb06464.x
- Feeney, D. A. (1997). From laboratory to clinic: Noradrenergic enhancement of physical therapy for stroke or trauma patients. *Advances in Neurology*, *73*, 383-394.
- Feeney, D. & Baron, J. (1986). Diaschisis. *Stroke*, *17*, 817-830.
- Fernaesus, S. & Almkvist, O. (1998). Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology*, *20*, 137-143. doi: 10.1076/jcen.20.2.137.1170
- Fernaesus, S., Ostberg, P., Hellstrom, A., & Wahlund, L. (2008). Cut the coda: Early fluency intervals predict diagnoses. *Cortex*, *44*, 161-169. doi: 10.1016/j.cortex.2006.04.002
- Fink, T. E. & Weisberg, R. W. (1981). The use of phonemic information to solve anagrams. *Memory and Cognition*, *9*, 402-410. doi: 10.3758/BF03197566
- Folstein, M.F, Folstein, S.E, & Mc Hugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.
- Frangakis, C. E. & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*, 21-29. doi 10.1111/j.0006-341X.2002.00021.x
- Fregni, F., Boggio, P. S., Valle, A. C., Rocha, R. R., Duarte, J. Ferreira, M. J. L.....Pascual-Leone, A. (2006). A sham-controlled trial of a 5-day course of repetitive transcranial magnetic stimulation of the unaffected hemisphere in stroke patients. *Stroke*, *37*, 2115-2122. doi: 10.1161/01.STR.0000231390.58967.6b
- Frerichs, R. J., & Tuokko, H. A. (2005). A comparison of methods for measuring cognitive change in older adults. *Archives of Clinical Neuropsychology*, *20*, 321-333.  
doi:10.1016/j.acn.2004.08.002

- Gallacher, K. I., Batty, G. D., McLean, G., Mercer, S. W., Guthrie, B., May, C. R.,... Mair, F. S. (2014). Stroke, multimorbidity and polypharmacy in a nationally representative sample of 1,424,378 patients in Scotland: Implications for treatment burden. *BMC Medicine*, *12*, 151. doi: 10.1186/s12916-014-0151-0
- Gathercole, S. E., Dunning, D. I., & Holmes, J. (2012). Cogmed training: Let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition*, *1*, 201-203. doi: 10.1016/j.jarmac.2012.07.007
- Golden, C. (1978). *Stroop color and word test*. Illinois: Stoelting Company
- Goodglass, H. & Kaplan, E. (1972). *Assessment of Aphasia and Related Disorders*. Philadelphia: Lea and Febiger.
- Goodglass, H. & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination*. Philadelphia, PA: Lea & Febiger.
- Grace, J., Nadler, J. D., White, D. A., Guilmette, T. J., Giuliano, A. J., Monsch, A. U., & Snow, M. G. (1995). Folstein vs Modified Mini-Mental State Examination in geriatric stroke stability, validity, and screening utility. *Archives of Neurology*, *52*, 477-484. doi: 10.1001/archneur.1995.00540290067019
- Grandmaison, E. & Simard, M. (2003). A critical review of memory stimulation programs in Alzheimers disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, *15*, 130-144. doi: 10.1176/appi.neuropsych.15.2.130
- Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*, 756-772. doi: 10.1007/s00426-013-0535-3
- Hachinski, V. (1999) Post-stroke depression, not to be underestimated. *Lancet*, *353* (9166), 1728. DOI: 10.1016/S0140-6736(99)00139-7.
- Hacke, W., Kaste, M., Bluhmki, E., Brozman, M., Davalos, A., Guidetti, D.....Toni, D (2008). Thrombolysis with alteplase 3 to 4.5 hours after acute Ischaemic stroke. *New England Journal of Medicine*, *359*, 1317-1329. doi: 101056/NEJMoa0804656

- Hales, J. & Brewer, J. (2012). The path to memory is guided by strategy: distinct networks are engaged in associative encoding under visual and verbal strategy and influence memory performance in healthy and impaired individuals. *Journal of Cognitive Neuroscience*, *24*, 1398-1411. doi: 10.1162/jocn\_a\_00220
- Hammar, Å., & Årdal, G. (2012). Effortful information processing in patients with major depression —A 10-year follow-up study. *Psychiatry Research*, *198*, 420-423. doi: 10.1016/j.psychres.2011.11.020
- Harris-Love, M. L. & Cohen, L. G. (2006). Noninvasive cortical stimulation in neurorehabilitation: A review. *Archives of Physical Medicine and Rehabilitation*, *87*, 84-93. doi: 10.1016/j.apmr.2006.08.330
- Hatano, S. (1976). Variability of the diagnosis of stroke by clinical judgement and by a scoring method. *Bulletin of the World Health Organization*, *54*, 533-540.
- Heiss, W. D., Huber, M., Fink, G. R., Herholz K., Pietrzyk, U., Wagner, R., & Wienhard, K. (1992). Progressive derangement of periinfarct viable tissue in Ischaemic stroke. *Journal of Cerebral Blood Flow & Metabolism*, *12*, 193-203. doi: 10.1038/jcbfm.1992.29
- Herrmann, M., Ehlis, A., & Fallgatter, A. (2003). Frontal activation during a verbal fluency task as measured by near-infrared spectroscopy. *Brain Research Bulletin*, *61*, 51-56. doi: 10.1016/S0361-9230(03)00066-2
- Hier, D. B., Mondlock, J., & Caplan, L. R. (1983). Recovery of behavioural abnormalities after right hemisphere stroke. *Neurology*, *33*, 345-350.
- Ho, A., Sahakian, B., Robbins, T., Barker, R., Rosser, A., & Hodges, J. (2002). Verbal fluency in Huntington's disease: A longitudinal analysis of phonemic and semantic clustering and switching. *Neuropsychologia*, *40*, 1277-1284. doi: 10.1016/S0028-3932(01)00217-2
- Hochstenbach, J., Mulder, T., Van Limbeek, J., Donders, R., & Schoonderwaldt, H. (1998). Cognitive decline following stroke: A comprehensive study of cognitive decline following

stroke. *Journal of Clinical and Experimental Neuropsychology*, *20*, 503-517. doi:  
10.1076/jcen.20.4.503.1471

Hochstenbach, J.B., Den Otter, R., & Mulder, T.W. (2003). Cognitive recovery after stroke: A 2-year follow-up. *Archives of Physical Medicine and Rehabilitation*, *84*, 1499-1504. doi:  
10.1016/S0003-9993(03)00370-8

Hoffman, P., Jefferies, E., Ehsan, S, Jones, R., & Lambon Ralph, M. (2012). How does linguistic knowledge contribute to short-term memory? Contrasting effects of impaired semantic knowledge and executive control. *Aphasiology*, *26*, 383-403. doi:  
10.1080/02687038.2011.581798

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9 – F15. doi: 10.1111/j.1467-7687.2009.00848.x

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science*, *16*(12), 927-931. doi: 10.1111/j.1467-9280.2005.01638.x

Howe, M. L., Garner, S. R., & Patel, M. (2013). Positive consequences of false memories. *Behavioural Sciences and the Law*, *31*, 652-665. doi: 10.1002/bsl.2078

Howe, M. L., Garner, S. R., Dewhurst, S. A., & Ball, L. J. (2010). Can false memories prime problem solutions? *Cognition*, *117*(2), 176-181. doi: 10.1016/j.cognition.2010.08.009

Howe, M. L., Threadgold, E., Norbury, J., Garner, S., & Ball, L. J. (2013). Priming children's and adults' analogical problem solutions with true and false memories. *Journal of Experimental Child Psychology*, *116*(1), 96-103. doi: 10.1016/j.jecp.2013.03.006

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*, 201-206. doi:  
10.1016/j.neuron.2004.12.033

- Humm, D. A., Kozlowski, D. A., James, D. C., Gotts, J. E., & Schallert, T. (1998). Use-dependent exacerbation of brain damage occurs during an early post-lesion vulnerable period. *Brain Research*, 783, 268-292. doi: 10.1016/S0006-8993(97)01356-5
- Humphreys, M. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, 4, 221-232. doi: 10.3758/BF03213167
- Hurford, R., Charidimou, A., Fox, Z., Cipolotti, L., & Werring, D. J. (2013). Domain specific trends in cognitive impairment after acute ischaemic stroke. *Journal of Neurology*, 260, 237-241. doi: 10.1007/s00415-012-6625-0
- Indredavik, B., Bakke, F., Slordahl, S. A., Rokseth, R., & Haheim, L. L. (1999). Treatment in a combined acute and rehabilitation stroke unit: Which aspects are most important? *Stroke*, 30, 917-923.
- Ivnik, R. J., Malec, J. F., Smith, G. E., & Tangalos, E. G. (1996). Neuropsychological test norms above age 55. COWAT, BNT, MAE token, WRAT-R reading, AMNART, Stroop, TMT, and JLO. *The Clinical Neuropsychologist*, 10, 262-278. doi: 10.1080/13854049608406689
- Jaeggi, S. Buschkuhl, M., Jonides, J. & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108, 10081-10086. doi: 10.1073/pnas.1103228108
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6829 – 6833. doi: 10.1073/pnas.0801268105
- Jaeggi, S., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory and Cognition*, 42, 464-480. doi: 10.3758/s13421-013-0364-z
- Jaillard, A., Naegele, B., Trabucco-Miguel, S., LeBas, J.F., & Hommel, M. (2009). Hidden dysfunctioning in subacute stroke. *Stroke*, 40, 2473-2479. doi: 10.1161/STROKEAHA.108.541144

- Johansson, B. B. (2004). Functional and cellular effects of environmental enrichment after experimental brain infarcts. *Restorative Neurology and Neuroscience*, *22*, 163-74.
- Jones, S., Laukka, E. J., & Baackman, L. (2006). Differential verbal fluency deficits in the preclinical stages of Alzheimer's disease and vascular dementia. *Cortex*, *42*, 347-355. doi: 10.1016/S0010-9452(08)70361-7
- Jung, W. H., Kim, S. N., Lee, T. Y., Jang, J. H., Choi, C. H., Kang, D. H., & Kwon, J. S. (2013). Exploring the brains of Baduk (Go) experts: Gray matter morphometry, resting-state functional connectivity, and graph theoretical analysis. *Frontiers in Human Neuroscience*, *633*. doi: 10.3389/fnhum.2013.00633
- Kane, M. J. & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47-70. doi: 10.1037/0096-3445.132.1.47
- Kane, M. J., Conway, A. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. in A. A. Conway, C. Jarrold, M. J. Kane, J. N. Towse (Eds.) , *Variation in working memory* (pp. 21-46). New York, NY, US: Oxford University Press.
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*, 978– 990. doi: 10.1111/j.1467-7687.2009.00846.x
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*, 2027-2037. doi: 10.1177/0956797614548725
- Katsman, D. T., Zheng, J., Spinelli, K., & Carmichael, S. T. (2003). Tissue microenvironments within functional cortical subdivisions adjacent to focal stroke. *Journal of Cerebral Blood Flow Metabolism*, *23*(9), 997-1009.

- Kertesz, A. (1982). *Western Aphasia Battery*. San Antonio: The Psychological Corporation.
- Kertesz, A. (1988). Is there a need for standardized aphasia tests? Why, how, what and when to test aphasics. *Aphasiology*, 2, 313-318. doi: 10.1080/02687038808248931
- Kertesz, A., & McCabe, P. (1977). Recovery patterns and prognosis in aphasia. *Brain: A Journal of Neurology*, 100, 1-18. doi: 10.1093/brain/100.1.1
- Khedr, E. M., Ahmed, M. A., Fathy, N., & Rothwell, J. C. (2005). Therapeutic trial of repetitive transcranial magnetic stimulation after acute Ischaemic stroke. *Neurology*, 65, 466-468. doi: 10.1212/01.wnl.0000173067.84247.36
- Kleider, H. M., Parrott, D. J., & King, T. Z. (2010). Shooting behaviour: How working memory and negative emotionality influence police officer shoot decisions. *Applied Cognitive Psychology*, 24(5), 707-717. doi: 10.1002/acp.1580
- Kleim, J. (2011). Neuralplasticity and neurorehabilitation: teaching the new brain old tricks. *Journal of Communication Disorders*, 44, 521-528. doi: 10.1016/j.jcomdis.2011.04.006
- Kleim, J., Barbay, S., Cooper, N. R., Hogg, T. R., Reidel, C. N., Remple, M. S., & Nudo, R. J. (2002). Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. *Neurobiology of Learning and Memory*, 77, 63-77. doi: 10.1006/nlme.2000.4004
- Kleim, J., Lussnig, E., Schwarz, E. R., Comery, T. A., & Greenough, W. T. (1996). Synaptogenesis and FOS expression in the motor cortex of the adult rat after motor skill learning. *Journal of Neuroscience*, 16, 4529-4535.
- Kleim, J., Pipitone, M. A., Czerlanis, C., & Greenough, W. T. (1998). Structural stability within the Lateral Cerebellar Nucleus of the rat following complex motor learning. *Neurobiology of Learning and Memory*, 69, 290-306. doi: 10.1006/nlme.1998.3828
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317-324. doi: 10.1016/j.tics.2010.05.002
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K.,.....Westerberg, H. (2005). Computerized training of working memory in children with ADHD – A

randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 177-186. doi: 10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24, 781-791. doi: 10.1076/jcen.24.6.781.8395

Kolb, B., & Wishaw, I. Q. (2009) *Fundamentals of Human Neuropsychology* (6th edition). New York: Worth.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K.P, Boxer, A. L... Widmeyer, M. (2014). NIH EXAMINER: conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society*, 20, 9-11. doi: 10.1017/S1355617713001094

Larsson, M. U., Almkvist, O., Luszcz, M. A., & Wahlin, T. (2008). Phonemic fluency deficits in asymptomatic gene carriers for Huntington's disease. *Neuropsychology*, 22, 596-605. doi: 10.1037/0894-4105.22.5596

Lees, K., Bluhmki, E., von Kummer, R., Brott, T. G., Toni, D., Grotta, J. C.....Hacke, W. (2010). Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *The Lancet*, 375, 1695-1703. doi: 10.1016/S0140-6736(10)60491-6

Lerner, I., Bentin, S., & Shriki, O. (2012). Spreading activation in an attractor network with latching dynamics: Automatic semantic priming revisited. *Cognitive Science*, 36, 1339-1382. doi: 10.1111/cogs.12007

Lezak, M. D., Howieson, D. B., & Loring, D. (2004). *Neuropsychological assessment* (4<sup>th</sup> edition). New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th edition). UK: Oxford University Press.

- Lexell, J. E. & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation. *American Journal of Physical Medicine & Rehabilitation*, *84*, 719-723. doi: 10.1097/01.phm.0000176452.17771.20
- Liman, T. G., Heuschmann, P. U., Endres, M., Flöel, A., Schwab, S., & Kolominsky-Rabas, P. L. (2011). Changes in cognitive function over 3 years after first-ever stroke and predictors of cognitive impairment and long-term cognitive stability: The Erlangen Stroke Project. *Dementia and Geriatric Cognitive Disorders*, *31*(4), 291-299. doi: 10.1159/000327358
- Lindley, R. J. (2008). Stroke rehabilitation. *Brain Impairment*, *9*, 97-102. doi: 10.1375/brim.9.2.97
- Loetscher, T. & Lincoln, N. B. (2013). Cognitive rehabilitation of attention deficits following stroke. *The Cochrane database of systematic reviews*, *5*, CD002842.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Madureira, S., Guerreiro, M., & Ferro, J. M. (2001). Dementia and cognitive impairment three months after stroke. *European Journal of Neurology*, *8*, 621-627. doi: 10.1046/j.1468-1331.2001.00332.x
- Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus*, *16*, 1091-1101. doi: 10.1002/hipo.20233
- Marin, F., Roldan, V., Climent, V., Garcia, A., Marco, P., & Lip, G. (2003). Is thrombogenesis in atrial fibrillation related to matrix metalloproteinase-1 and its inhibitor, TIMP-1? *Stroke*, *34*, 1181-1186.
- Marshall, R., Perera, G., Lazar, R., Krakauer, J. W., Constantine, R. C., & DeLaPaz, R. L. (2000). Evolution of cortical activation during recovery from corticospinal tract infarction. *Stroke* *31*, 656-661.

- Maulden, S. A., Gassaway, J., Horn, S. D., Smout, R. J., & DeJong, G. (2005). Timing of initiation of rehabilitation after stroke. *Archives of Physical Medicine and Rehabilitation*, *86*, 34-40. doi: 10.1016/j.apmr.2005.08.119
- McEvoy, C. L. & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, *95*, 581-634. doi: 10.2307/1422189
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220-232. doi: 10.1037/hoo48850
- Meinz, E. J. & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science*, *21*(7), 914-919. doi: 10.1177/0956797610373933
- Meinzer, M., Elbert, T., Djundja, D., Taub, E., & Rockstroh, B. (2007). Extending the Constraint-Induced Movement Therapy (CMIT) approach to cognitive functions: Constraint-Induced Aphasia Therapy (CIAT) of chronic aphasia. *NeuroRehabilitation*, *22*, 311-318.
- Melby-Lervåg, M. & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270-291. doi: 10.1037/a0028228
- Meyer, B.-U., Roricht, S., von Einsiedel, H. G., Kruggel, F., & Weindl, A. (1995). Inhibitory and excitatory interhemispheric transfers between motor cortical areas in normal humans and patients with abnormalities to the corpus callosum. *Brain*, *118*, 429-440.
- Mezzacappa, E. & Buckner, J. C. (2010). Working memory training for children with attention problems or hyperactivity: A school-based pilot study. *School Mental Health*, *2*, 202-208. doi: 10.1007/s12310-010-9030-9
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi: 10.1006/cogp.1999.0734

- Morrison, A. B. & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46-60. doi: 10.3758/s13423-010-0034-0
- Muncer, S. & Knight, D. (2011). The syllable effect in anagram solution: Unrecognised evidence from past studies. *Journal of Psycholinguistic Research*, *40*, 111-188. doi: 10.1007/s10936-010-9159-6
- Murphy, T. H. & Corbett, D. (2009). Plasticity during stroke recovery: From synapse to behaviour. *Nature Reviews Neuroscience*, *10*(12), 861-872. doi: 10.1038/nrn2735
- Musicco, M., Emberti, L., Nappi, G., & Caltagirone, C. (2003). Early and long-term outcome of rehabilitation in stroke patients: The role of patient characteristics, time of initiation, and duration of interventions. *Archives of Physical Medicine and Rehabilitation*, *84*, 551-558. doi: 10.1053/apmr.2003.50084
- Nasreddine, Z. S., Phillips, N. A., Bedivian, V., Charvonneau, S., Whitehead, V., Collins, I.,...Chertkow, H. (2005). The Montreal Cognitive Assessment MoCA: brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*, 695-699. doi: 10.1111/j.1532-5415.2005.53221.x
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. (1998). *Word association, rhyme and word fragment norms*. <http://w3.usf.edu/FreeAssociation>
- Nhan, H., Barquist, K., Bell, K., Esselman, P., Adderson, I. R., & Cramer, S. C. (2003). A serial study of brain function after stroke. *Stroke*, *34*, 253.
- Nordvik, J. E., Schanke, A. K., Walhovd, K., Fjell, A., Grydeland, H., & Landro, N., J. (2012). Exploring the relationship between white matter microstructure and working memory functioning following stroke: A single case study of computerised cognitive training. *Neurocase*, *18*, 139-151. doi: 10.1080/13554794.568501

- Novick, L. R. & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *56A*, 351-382. doi: 10.1080/02724980244000288
- Novick, L. R. & Sherman, S. J. (2008). The effects of superficial and structural information on online problem solving for good versus poor anagram solvers. *The Quarterly Journal of Experimental Psychology*, *61*, 1098-1120. doi: 10.1080/17470210701449936
- Nudo, R. J. (2006). Mechanisms for recovery of motor function following cortical damage. *Current Opinion in Neurobiology*, *16*, 638-644.
- Nudo, R. J. & Milliken, G. W. (1996). Reorganization of movement representations in primary motor cortex following focal Ischaemic infarcts in adult squirrel monkeys. *Journal of Neurophysiology*, *75*, 2144-2149.
- Nys, G. M. S., Van Zandvoort, M. J. E., de Kort, P. L. M., Jansen, B. P. W., Kappelle, L. J., & de Haan, E. H. (2005). Restrictions of the Mini-Mental State Examination in acute stroke. *Archives of Clinical Neuropsychology*, *20*, 623-629. doi: 10.1016/j.acn.2005.04.001
- Nys, G. M. S., Van Zandvoort, M. J. E., de Kort, P. L. M., Jansen, B. P. W., Van Der Worp, Kappelle, L. J., & de Haan, E. H. (2005). Domain-specific cognitive recovery after first-ever stroke: A follow-up study of 111 cases. *Journal of the International Neuropsychological Society*, *11*, 795-806. doi: 10.1017/S1355617705050952
- Nys, G. M. S., Van Zandvoort, M. J. E., de Kort, P. L. M., Jansen, B. P. W., de Haan, E. H., & Kappelle, L. J. (2007). Cognitive disorders in acute stroke: Prevalence and clinical determinants. *Cerebrovascular Diseases*, *23*, 408-416. doi: 10.1159/000101464
- O'Sullivan, M., Morris, R. G., & Markus, H. S. (2005). Brief cognitive assessment for patients with cerebral small vessel disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *76*, 1140-1145. doi: 10.1136/jnnp.2004.045963

- Ostrosky-Solis, F., Ardila, A., & Rosselli, M. (1999). NEUROPSI: A brief neuropsychological test battery in Spanish with norms by age and educational level. *Journal of the International Neuropsychological Society*, *5*, 413-133. doi: 10.1017/S1355617799555045
- Ostrosky-Solis, F., Ramirez, M., Lozano, A., Picasso, H., & Velez, A. (2004). Culture or education? Neuropsychological test performance of a Maya indigenous population. *International Journal of Psychology*, *39*, 36-46. doi: 10.1080/00207590344000277
- Palmer, E., Ashby, P., & Hajek, V. (1992). Ipsilateral fast corticospinal pathways do not account for recovery in stroke. *Annals of Neurology*, *32*, 519–525. doi: 10.1002/ana.410320407
- Pascual-Leone, A., Dang, N., Cohen, L. G., Brasil-Neto, J. P., Cammarota, A., & Hallett, M. (1995). Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. *Journal of Neurophysiology*, *74*, 1037-1045.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of human knowledge in the brain. *Nature Reviews Neuroscience*, *8*, 976-987.
- Patel, M., Coshall, C., Rudd, A. G., & Wolfe, C. D A. (2003). Natural history of cognitive impairment after stroke and factors associated with its recovery. *Clinical Rehabilitation*, *17*, 158-166. doi: 10.1191/0269215503cr596oa
- Perez, M. A., Lungholt, B. K., Nyborg, K., & Neilsen, J. B. (2004). Motor skills training induces changes in the excitability of the leg cortical area in healthy humans. *Experimental Brain Research*, *159*, 197-205. doi: 10.1007/s00221-004-1947-5
- Pfeiffer, H. V., Løkkegaard, A., Zoetmulder, M., Friberg, L., & Werdelin, L. (2014). Cognitive impairment in early-stage non-demented Parkinson's disease patients. *Acta Neurologica Scandinavica*, *129*, 307-318. doi: 10.1111/ane.12189
- Planton, M., Peiffer, S., Albucher, J. F., Barbeau, E. J., Tardy, J., Pastor, J..... Pariente, J. (2012). Neuropsychological outcome after a first symptomatic ischaemic stroke with “good recovery”. *European Journal of Neurology*, *19*, 212-219. doi: 10.1111/j.1468-1331.2011.03450.x

- Ponsford, J., Sloan, S., & Snow, P. (2013). *Traumatic brain injury: Rehabilitation for everyday adaptive living*. Hove, United Kingdom: Psychology Press.
- Poulin, V., Korner-Bitensky, N., Dawson, D. R., & Bherer, L. (2012). Efficacy of executive function interventions after stroke: A systemic review. *Topics in Stroke Rehabilitation, 19*, 158-171. doi: 10.1310/tsr1902-158
- Pustokhanova, L., & Morozova, E. (2013). Cognitive impairment and hypothymia in post-stroke patients. *Journal of the Neurological Sciences, 325*(1-2), 43-45. doi: 10.1016/j.jns.2012.11.013
- Rabipour, S. & Raz, A. (2012). Training the brain: Fact and fad in cognitive and behavioural remediation. *Brain and Cognition, 79*, 159 – 179. doi: 10.1016/j.bandc.2012.02.006
- Raboutet, C., Sauzeon, H., Corsini, M., Rodrigues, J., Langevin, S., & N’Kaoua, B. (2010). Performance on a semantic verbal fluency task across time: Dissociation between clustering, switching, and categorical exploitation processes. *Journal of Clinical and Experimental Neuropsychology, 32*, 268-280. doi: 10.1080/13803390902984464
- Reischies, F. M. & Berghöfer, A. (1995). Saccadic tracking test: Normal data and reliability. *Journal of Clinical Psychology, 51*(2), 252-257. doi: 10.1002/1097-4679(199503)51:2<252::AID-JCLP2270510215>3.0.CO;2-S
- Robertson, I. H. & Murre, J. J. (1999). Rehabilitation of brain damage: Brain plasticity and principles of guided recovery. *Psychological Bulletin, 125*, 544-575. doi: 10.1037/0033-2909.125.5.544
- Roediger, H. L. & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814. doi: 10.1037/0278-7393.21.4.803
- Rogers, W. (2010). Parallel Forms Reliability. In Neil J. Salkind (Ed.), *Encyclopedia of research design*. (pp. 996-998). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412961288.n301

- Sachdev, P. S., Brodaty, H., Valenzuela, M. J., Lorentz, L. M., & Koschera, A. (2004a). Progression of cognitive impairment in stroke patients. *Neurology*, *63*, 1618-1623.
- Sachdev, P. S., Brodaty, H., Valenzuela, M. J., Lorentz, L. M., Looi, J. C. L., Wen, W., & Zagami, A. S. (2004b). The neuropsychological profile of vascular cognitive impairment in stroke and TIA patients. *Neurology*, *62*, 912-919.
- Schallert, T. B., Leasure, J. L., & Kolb, B. (2000). Experience-associated structural events, subependymal cellular proliferative activity, and functional recovery after injury to the central nervous system. *Journal of Cerebral Blood Flow and Metabolism*, *20*, 1513-1528. doi: 10.1097/00004647-200011000-00001
- Schlaug, G., Jancke, L., Huang, Y., Staiger, J., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*, 1047-1055. doi: 10.1016/0028-3932(95)00045-5
- Schneider, S. G. & Asarnow, R. F. (1987). A comparison of cognitive/neuropsychological impairments of nonretarded autistic and schizophrenic children. *Journal of Abnormal Child Psychology*, *15*, 29-46. doi: 10.1007/BF00916464
- Schramke, C. J., Stowe, R. M., Ratcliff, G., Goldstein, G., & Condray, R. (1998). Poststroke depression and anxiety: Different assessment methods result in variations in incidence and severity estimates. *Journal of Clinical and Experimental Neuropsychology*, *20*, 723-737. doi: 10.1076/jcen.20.5.723.1117
- Seitz, R., Hoflich, P., Binkofski, F., Tellmann, L., Herzog, H., & Freund, H. J. (1998). Role of the premotor cortex in recovery from middle cerebral artery infarction. *Archives of Neurology*, *55*, 1081-1088. doi: 10.1001/archneur.55.8.1081
- Senior, G., Kelly, M., & Salzman, L. (1999). Clinical utility of the Rey tangled lines test. *Archives of Clinical Neuropsychology*, *14*, 719-719. doi: 10.1093/arclin/14.8.719
- Seyal, M., Ro., T., & Rafal, R. (1995). Increased sensitivity to ipsilateral cutaneous stimuli following transcranial magnetic stimulation of the parietal lobe. *Annals of Neurology*, *38*, 264-267. doi: 10.1002/ana.410380221

- Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition, 1*, 185-193. doi: 10.1016/j.jarmac.2012.06.003
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica, 50*, 245-276. doi: 10.5334/pb-50-3-4-245
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*, 628-654. doi: 10.1037/a0027473
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi: 10.1037/0033-2909.86.2.420
- Spiers, H. J. & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *NeuroImage, 31*, 1826-1840. doi: 10.1016/j.neuroimage.2006.01.037
- Srikanth, V. K., Thrift, A. G., Saling, M. M., Anderson, J. F., Dewey, H.M., MacDonnell, R. A., & Donnan, G. A. (2003). A community-based prospective study of nonaphasic English-speaking survivors. *Stroke, 34*, 1136-1143. doi: 10.1161/01.STR.0000069161.35736.39
- Stablum, F., Umilta, C., Mogentale, C., Carlan, M., & Guerrini, C. (2000). Rehabilitation of executive deficits in closed head injury and anterior communicating artery aneurysm patients. *Psychological Research, 63*, 265-278. doi: 10.1007/s004269900002
- Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory and Cognition, 27*, 494-500. doi: 10.3758/BF03211543
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests*. Oxford, UK: Oxford University Press.
- Stuart, E.A. & Rubin, D. B. (2007). Best Practices in Quasi-Experimental Designs: Matching methods for causal inference. Chapter 11 (pp. 155-176) in *Best Practices in Quantitative Social Science*. J. Osborne (Ed.). Thousand Oaks, CA: Sage Publications.

- Svensson, P., Romaniello, A., Arendt-Neilsen, L., & Sessle, B. J. (2003). Plasticity in corticomotor control of the human tongue musculature induced by tongue-task training. *Experimental Brain Research, 152*, 42-51. doi: 10.1007/s00221-003-1517-2
- Tatemichi, T. K., Desmond, D. W., Stern, Y., Paik, M., Sano, M., & Bagiella, E. (1994). Cognitive impairment after stroke: frequency, patterns, and relationship to functional abilities. *Journal of Neurology, Neurosurgery and Psychiatry, 57*, 202-207. doi: 10.1136/jnnp.57.2.202
- Taub, E. & Wolf, S. L. (1997). Constraint induced movement techniques to facilitate upper extremity use in stroke patients. *Topics in Stroke Rehabilitation, 3*, 38-61.
- Taub, E., Miller, N. E. Novack, T. A., Cook, E. W., Fleming, W. C., Nepomuceno, C. S. Connell, J. S., & Crago, J. E. (1993). Technique to improve chronic motor deficit after stroke. *Archives of Physical Medicine and Rehabilitation, 74*, 347-354.
- Taylor, G. H. & Broomfield, N. M. (2013). Cognitive assessment and rehabilitation pathway for stroke (CARPS). *Topics in Stroke Rehabilitation, 20*, 270-282. doi: 10.1310/tsr2003-270
- Tehan, G. (2010). Associative relatedness enhances recall and produces false memories in immediate serial recall. *Canadian Journal of Experimental Psychology, 64* (4), 266-272. doi: 10.1037/a0021375
- Testa, J., Troster, A., Fields, J., Gleason, A., Salmon, D., & Beatty, W. (1998). Semantic fluency performance of patients with cortical and subcortical neurodegenerative diseases. *Aging, Neuropsychology and Cognition, 5*, 208-214. doi: 10.1076/anec.5.3.203.612
- Tham, W., Auchus, A. P., Thong, M., Goh, M-L., Chang, H-M, Wong, M-C., & Chen, C. P. (2002) Progression of cognitive impairment after stroke: One year results from a longitudinal study of Singaporean stroke patient. *Journal of the Neurological Sciences, 203–204*, 49-52. doi: 10.1016/S0022-510X(02)00260-5.
- The National Institute of Neurological Disorders and Stroke. (1995). Tissue plasminogen activator for acute Ischaemic stroke. *New England Journal of Medicine, 333*, 1581-1588. doi: 10.1056/NEJM199512143332401

- Tombaugh, T. & McIntyre, N. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40*, 922-935.
- Tombaugh, T., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14* (2), 167-177. doi: 10.1016/S0887-6177(97)00095-4
- Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*, 499-504. doi: 10.1016/S0028-3932(97)00152-8
- Tse, C-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874–891. doi: 10.1080/09658210903376250
- Tulving, E. (1985). How Many Memory Systems Are There? *American Psychologist*, *40*, 385-398. doi: 10.1037/0003-066X.40.4.385
- Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154. doi: 10.1016/0749-596X(89)90040-5
- Unsworth, N. & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038-1066. doi: 10.1037/0033-2909.133.6.1038
- Unsworth, N. & Engle, R. W. (2007b). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104-132. doi: 10.1037/0033-295X.114.1.104
- Unsworth, N. & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, *17*(2), 200-205. doi: 10.3758/PBR.17.2.200
- Vallat, C., Azouvi, P., Hardisson, H., Meffert, R., Tessier, C., & Pradat-Diehl, P. (2005). Rehabilitation of verbal working memory after left hemisphere stroke. *Brain Injury*, *19*, 1157-1164. doi: 10.1080/02699050500110595

- van der Zijden, J. P., Wu, O., van der Toorn, A., Roeking, T. P., Bleys, R. L., & Dijkhuizen, R. M. (2007). Changes in neuronal connectivity after stroke in rats as studied by serial manganese-enhanced MRI. *NeuroImage*, *34*, 1650–1657. doi: 10.1016/j.neuroimage.2006.11.001
- van Geldorp, B., Kessels, R., & Hendriks, M. (2013). Single-item and associative working memory in stroke patients. *Behavioural Neurology*, *26*, 199-201 doi: 10.3233/BEN-2012-129010
- Van Zandvoort, M.J.E., Kessels, R.P.C., Nys, G.M.S., De Haan, E.H.F., & Kappelle, L.J. (2005) Early neuropsychological evaluation in patients with ischaemic stroke provides valid information. *Clinical Neurology and Neurosurgery*, *107*, 385-392. doi: 10.1016/j.clineuro.2004.10.012
- Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: How to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1322-1337. doi: 10.1037/0278-7393.30.6.1322
- Wade, D. T., Langton-Hewer, R, Wood, V. A., Skillbeck, C. E., & Ismail, H. M. (1983). The hemiplegic arm after stroke: Measurement and recovery. *Journal of Neurology, Neurosurgery and Psychiatry*, *46*, 521-524
- Wall, P. D. & Egger M. D. (1971). Formation of new connections in adult rat brains following partial deafferentation. *Nature*, *232*, 542–545.
- Wall, J. T. & Kass, J. H. (1985). Cortical reorganisation and sensory recovery following nerved damage and regeneration. In C. W. Cotman (Ed.) *Synaptic plasticity* (pp.231-259). New York: Guilford Press.
- Ward, N. S. & Cohen, L. G. (2004). Mechanisms underlying recovery of motor function after stroke. *Archives of Neurology*, *61*, 1844–1848. doi: 10.1001/archneur.61.12.1844
- Ward, N., Brown, M., Thompson, A., & Frackowiak, R. (2003). Neural correlates of motor recovery after stroke: a longitudinal fMRI study. *Brain*, *126*, 2476–2496. doi: 10.1093/brain/awg245
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition*. San Antonio, TX: Pearson.

- Weingartner, H., Eckardt, M., Grafman, J., Molchan, S., Putnam, K., Rawlings, R., & Sunderland, T. (1993). The Effects of repetition on memory performance in cognitively impaired patients. *Neuropsychology, 7*, 385-395. doi: 10.1037/0894-4105.7.3.385.
- Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Ostensson, M. L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke – A pilot study. *Brain Injury, 21*, 21-19. doi: 10.1080/02699050601148726
- Wieloch, T., & Nikolich, K. (2006). Mechanisms of neural plasticity following brain injury. *Current Opinion in Neurobiology, 16*, 258–264. doi: 10.1016/j.conb.2006.05.011
- Will, B. & Kelche, C. (1992). Environmental approaches to recovery of function from brain damage: A review of animal studies (1981 to 1991). In F. D. Rose, D. A. Johnson (Eds.), *Recovery from brain damage: Reflections and directions*. (pp. 79-103). New York, NY, US: Plenum Press. doi: 10.1007/978-1-4615-3420-4\_5
- Wilson, B.A. (2009). *Memory rehabilitation*. New York, NY: The Guilford Press.
- Withers, G. S. & Greenough, W. T. (1989). Reach training selectively alters dendritic branching in subpopulations of layer II-III pyramids in rat motor-somatosensory forelimb cortex. *Neuropsychologia, 27*, 61-69. doi: 10.1016/0028-3932(89)90090-0
- Wolf, S. L., Winstein, C. J., Miller, J. P., Taub, E., Uswatte, G., Morris, D.....Nichols-Larsen, D. (2006). Effect of Constraint-Induced Movement Therapy on upper extremity function 3 to 9 months after stroke: The EXCITE randomized clinical trial. *The Journal of the American Medical Association, 296*, 2096-2104. doi: 10.1001/jama.296.17.2095
- Wolf, S. L., Winstein, C. J., Miller, J. P., Thompson, P. A., Taub, E., Uswatte, G.....Clark, P. C. (2008). Retention of upper limb function in stroke survivors who have received constraint-induced movement therapy: the EXCITE randomized trial. *The Lancet Neurology, 7*, 33-40. doi: 10.1016/S1474-4422(07)70294-6

Zec, R. F., Landreth, E. S., Fritz, S., Grames, E., Hasara, A., Fraizer, W.....Manyam, B. (1999). A comparison of phonemic, semantic, and alternating word fluency in Parkinson's disease. *Archives of Clinical Neuropsychology, 14*, 255-264. doi: 10.1016/S0887-6177(98)00008-0

## Appendix A

### Human Research Ethics Approval Forms

Human Research Ethics Committees ACU and The Prince Charles Hospital

Committee Approval Form

**Principal Investigator/Supervisor:** Associate Professor Anne Tolan- Brisbane Campus

**Co-Investigators:** Dr Kate Witteveen

**Student Researcher:** Hannah Tehan Brisbane Campus

**Ethics approval has been granted for the following project:**

Taining in prospective memory for those suffering the effects of Stroke

**for the period:** 16 December 2011 to 31 January 2014

**Human Research Ethics Committee (HREC) Register Number:** Q2011 73

The following standard conditions as stipulated in the *National Statement on Ethical Conduct in Research Involving Humans (2007)* apply:

- (i) that Principal Investigators / Supervisors provide, on the form supplied by the Human Research Ethics Committee, annual reports on matters such as:
  - security of records
  - compliance with approved consent procedures and documentation
  - compliance with special conditions, and
  
- (ii) that researchers report to the HREC immediately any matter that might affect the ethical acceptability of the protocol, such as:
  - proposed changes to the protocol
  - unforeseen circumstances or events
  - adverse effects on participant

The HREC will conduct an audit each year of all projects deemed to be of more than low risk. There will also be random audits of a sample of projects considered to be of negligible risk and low risk on all campuses each year.

Within one month of the conclusion of the project, researchers are required to complete a *Final Report Form* and submit it to the local Research Services Officer.

If the project continues for more than one year, researchers are required to complete an *Annual Progress Report Form* and submit it to the local Research Services Officer within one month of the anniversary date of the ethics approval.



Signed:

.....Date: 23 April 2010

(Research Services Officer, McAuley Campus)

---

Ethics Register Number : Q2011 73

Project Title : Training in prospective memory for those suffering the effects of Stroke.

Enquiries to: [RGOTPCH@health.qld.gov.au](mailto:RGOTPCH@health.qld.gov.au)  
 Office Ph: [Anne\\_Carle@health.qld.gov.au](mailto:Anne_Carle@health.qld.gov.au)  
 (07) 3139 4198  
 Our Ref: (07) 3139 4407  
 AC/JL/Governance Approval



14 November 2013

**Research Governance Office**  
 Metro North Hospital and Health Service  
 The Prince Charles Hospital  
 Administration Building, Lower Ground  
 Rode Road, Chermide QLD 4032

Ms Hannah Tehan  
 C/- Dr Mia Mariani  
 Rehabilitation Ward  
 The Prince Charles Hospital

Dear Ms Tehan,

**HREC reference number: HREC/12/QPCH/87**  
**SSA reference number: SSA/13/QPCH/250**  
**Project title: Retraining in Prospective Memory - RPM: Brain in Gear.**

Thank you for submitting an application for authorisation of this project. I am pleased to inform you that authorisation has been granted for this study to take place at the following site:

**The Prince Charles Hospital**

The following conditions apply to this research proposal. These are additional to those conditions imposed by the Human Research Ethics Committee that granted ethical approval.

1. Proposed amendments to the research protocol or conduct of the research which may affect the ethical acceptability of the project are to be submitted to the HREC for review. A copy of the HREC approval/rejection letter must be submitted to the RGO;
2. Proposed amendments to the research protocol or conduct of the research which only affects the ongoing site acceptability of the project, are to be submitted to the research governance officer;
3. Proposed amendments to the research protocol or conduct of the research which may affect both the going ethical acceptability of the project and the site acceptability of the project are to be submitted firstly to the HREC for review and then to the research governance officer after a HREC decision is made.

I am pleased to advise Governance approval of this research project. The documents reviewed and approved include:

<i>Document</i>	<i>Version</i>	<i>Date</i>
Site Specific Application (AU/3/7354113)		

Please complete the Notification of Commencement Form once commencement of this protocol has occurred at this site ([http://www.health.qld.gov.au/tpch/documents/form\\_notification.dot](http://www.health.qld.gov.au/tpch/documents/form_notification.dot)) and return to the office of the Human Research Ethics Committee.

Office	Postal	Phone
Research, Ethics & Governance Office The Prince Charles Hospital	Administration Building, Lower Ground Rode Road, Chermide Q 4032	(07) 3139 4407

On behalf of the Research, Ethics and Governance Unit, we wish you every success in your research project.

Yours sincerely

**Anne Carle**  
CEO Delegate  
Metro North Hospital and Health Service

## Appendix B

### Information Letter and Consent Instructions used in Experiments 4.1, 4.2. and 5.1

#### INFORMATION LETTER TO PARTICIPANTS

TITLE OF PROJECT: Brain Training in Prospective Memory

STAFF SUPERVISOR: A/Prof Anne Tolan

STUDENT RESEARCHERS: Hannah Tehan

PROGRAMME IN WHICH ENROLLED: Master of Psychology (Clinical)/PhD

Dear Participant,

The research I am doing for my PhD is aimed at providing and evaluating a cognitive training program for people who have suffered a stroke. In developing and evaluating this program we need participants to complete a series of cognitive tasks in order to understand the relationships among these tasks and how effective they may be in terms of brain training.

You are invited to participate in this study which will explore a training program designed to improve memory for future events. You will be given six activities:

- The card game snap
- Anagrams (word jumbles)
- Short-term Memory Task
- A colour reading task
- A word association task
- A verbal fluency task
- A visual maze task

Before completing each task you will be given an explanation of each task and instructions on how to complete it. I will also collect from you demographic details, age, gender, years of education and some questions about your current health status.

**Please Note:** These tasks usually take less than half an hour to complete; **However**, because this is a training program, and the idea is to practice these tasks. **The study requires participants to be available for half an hour a day, for six days across a two week period for the researcher to administer the training.** It is understood that this is a significant commitment and the researcher will be as flexible as possible to fit in with your schedule.

This training may benefit you in gaining some understanding into your overall memory ability, and the possibility of improving it. Furthermore, you will be kindly contributing to the pool of knowledge of prospective memory. At the end of the two week training program, you will also have the opportunity to see if/how your individual results have changed over the two week period!

Participation in this research study is voluntary. **You are able to withdraw at any stage without giving any reason. Refusal or withdrawal will not result in any disadvantage for you.**

There are no foreseeable risks in participating in this study, however, if you have concern about the topic in general or as a result of this study, it is encouraged that you seek information from the researcher or information and advice from your general practitioner.

The researchers will take every precaution to ensure confidentiality. All participants will be given a code and names will not be retained with the data. Individual participants will not be identified in any future presentation of the results; only group results will be reported. The group results will be reported in a thesis that may be published. Group results may also be provided to other researchers in an aggregated form that will not identify you in any way.

The Human Research Ethics Committee at Australian Catholic University has approved this study. In the event that you have any complaint or concern about the way you have been treated during the study, or if you have any query that the Supervisor and Student Researcher have not been able to satisfy, you may write to:

Chair, Human Research Ethics Committee  
C/o Research Services  
Australian Catholic University  
Brisbane Campus  
Po Box 456 – Virginia, QLD 4104  
TEL: 07 3623 7429 FAX: 07 3623 7328

Any complaint or concern will be treated in confidence and fully investigated. The participant will be informed of the outcome.

If you are willing to participate, please sign the attached consent forms. You should sign both copies of the consent form and keep one copy for your records and return the other copy to the staff supervisor or student researcher. Your support for the research project will be much appreciated.

**CONSENT FORM**

TITLE OF PROJECT: Brain Training in Prospective Memory

STAFF SUPERVISOR: A/Prof Anne Tolan

STUDENT RESEARCHER: Hannah Tehan

COURSE: Doctor of Philosophy

**Participants section**

I ..... (*the participant*) have read and understood the information in the information letter inviting participation to complete a series of cognitive tasks including:

- The card game snap
- Anagrams (word jumbles)
- Short-term Memory Task
- A colour reading task
- A word association task
- A verbal fluency task
- A visual maze task

Any questions I have asked have been answered to my satisfaction. I agree to attend a half an hour session for six sessions over a two week period. I realize that I can withdraw at any time without penalty.

I agree that research data collected for the study may be published or provided to other researchers in a form that does not identify me in any way. I am over 18 years of age.

NAME OF PARTICIPANT:

.....

SIGNATURE .....

DATE .....

**Appendix C:****Information Letter and Consent Instructions used in Case Studies****Participant Information Sheet**

<b>HREC No:</b>	HREC/12/QPCH/87
<b>Project Title:</b>	RPM: Brain in Gear
<b>Name of Researchers:</b>	Hannah Tehan & Dr Mia Mariani

**Title of Project**

Brain Training in Prospective Memory

**Invitation**

You are invited to participate in this study which aims to explore a training program designed to improve memory for future events.

**Background**

This study is aimed at providing and evaluating a cognitive training program for people who have suffered a stroke. In developing and evaluating this program we need participants to complete a series of cognitive tasks in order to understand the relationships among these tasks and how effective they may be in improving memory.

This is a training program, which means that participants are asked to complete the study every day over two weeks. There are three to five tasks every day and it should only take half an hour to complete. The tasks are:

- The card game snap
- Anagrams (word jumbles)
- Short-term Memory Task
- A colour reading task
- A word association task
- A verbal fluency task
- A visual maze task

Before you begin the program, there will be an explanation of each task and the chance to practice. The researcher will be with you when you complete the tasks, so any questions you may have along the way can be happily answered.

In addition to these tasks, the researcher would like to ask some more questions about you, such as your age, your years of education and some information about your health status.

**Benefits**

This training may benefit you in gaining some understanding into your overall memory ability, and the possibility of improving it. Furthermore, you will be kindly contributing to the pool of knowledge of prospective memory. At the end of the two week training program, you will also have the opportunity to see if/how your individual results have changed over the two week period!

**Risks and Side Effects**

There are no foreseeable risks in participating in this study, however, if you have concern about the topic in general or as a result of this study, it is encouraged that you seek information from the researcher or information and advice from your general practitioner.

**Confidentiality and Privacy**

The researchers will take every precaution to ensure confidentiality. All participants will be given a code and names will not be retained with the data. Individual participants will not be identified in any future presentation of the results; only group results will be reported. The group results will be reported in a thesis that may be published. Group results may also be provided to other researchers in an aggregated form that will not identify you in any way.

Participation in this research study is voluntary. You are able to withdraw at any stage without giving any reason. Refusal or withdrawal will not result in any disadvantage for you.

**Further Information**

Any questions regarding this project, before or after participating, should be directed to the Staff Supervisor, Dr Mia Mariani on 07 3139 5325 in the Department of Clinical Neuropsychology, The Prince Charles Hospital. Before commencing, you will have the opportunity to ask any questions about the project. You will also have the opportunity to discuss your participation and the project in general after completing the experiment.

**Independent Contact**

The Human Research Ethics Committee at The Prince Charles Hospital has approved this study. In the event that you have any complaint or concern about the way you have been treated during the study, or if you have any query that the Researchers have not been able to satisfy, you may write to or phone:

Mr Philip Lee  
The Prince Charles Hospital  
Metro North Health Service District  
Administration Building, Lower Ground  
Rode Road Chermside, QLD 4032  
PH: (07) 3139 4500

Any complaint or concern will be treated in confidence and fully investigated. The participant will be informed of the outcome.

If you are willing to participate, please sign the attached consent forms. Your participation in this study is much appreciated.



# Revocation of Consent Form – Participant

<b>HREC No:</b>	HREC/12/QPCH/87
<b>Project Title:</b>	RPM: Brain in Gear
<b>Name of Researchers:</b>	Hannah Tehan & Dr Mia Mariani

- I hereby wish to WITHDRAW my consent to participate in the research project described above and understand that such withdrawal WILL NOT jeopardise any treatment or my relationship with The Prince Charles Hospital Metro North Health Service District or The Australian Catholic University.

Participant's name (please print): .....

(Signature).....

DD / MMM / YYYY

Date: \_\_ / \_\_\_ / \_\_\_\_

---

This Revocation of Consent should be forwarded to:

Dr Mia Mariani  
Rehabilitation Unit  
The Prince Charles Hospital  
Metro North Health Service District  
Rode Road, Chermside, QLD 4032

---