The Development of Problem-Solving Abilities in Typical and Atypical Development

Joanne Sara Camp

Submitted for the degree of Doctor of Philosophy

Institute of Education, University of London

Abstract

Throughout our lives we engage in problem solving, which is thought to depend on executive functions (EFs) e.g., inhibition, shifting and working memory. Previous work has identified the need to consider these abilities in an everyday context. EF skills are known to be impaired in Williams syndrome (WS) and Down syndrome (DS). This thesis aims to investigate experimental and real-life problem solving in WS and DS, and how these groups use EF skills to solve problems, through experimental and questionnaire-based cross-syndrome comparisons.

Participants with WS and DS aged 12-24 years (Ns=20) and typically developing (TD) controls (N=56; nonverbal matched subset = 20) completed the Tower of London (TOL) problem-solving task and a battery of EF tests. In a separate study, parents (WS, DS, TD; total N=112) completed the BRIEF (Behavioral Rating Inventory of Executive Functioning) and a novel Problem-Solving Questionnaire.

The WS group, but not the DS group, scored more poorly on the TOL than the nonverbal-matched controls. In WS, developmental trajectory analysis indicated over-reliance on planning for TOL performance for low planning scores. For the DS group only speed of picture matching was associated with TOL performance, while more rule violations were exhibited than for the WS group. Questionnaire scores were poor for the WS group in relation to DS and TD groups. Asking for help for the DS group, and becoming emotional for the WS group, was related to reaching the solution. In general, associations between experimental and everyday measures were scarce.

It was concluded that: while EFs (planning, visuospatial working memory) were constraining factors for WS problem solving, alternative strategies were used by the DS group to reach the solution; real-life problem solving should be considered in its own right; and poor WS problem solving may be related to emotional difficulties.

I hereby declare that, except where explicit attribution is made,

the work presented in this thesis is entirely my own.

Word count (exclusive of list of references and appendices):

87622 words

Publications

In the course of undertaking the study of Williams syndrome for work towards this thesis, the following publications were produced.

Camp, J. S., Farran, E. & Karmiloff-Smith, A. (2012) 'Numeracy'. In E. Farran and A. Karmiloff-Smith (eds.) *Neurodevelopmental disorders across the lifespan: the neuroconstructivist approach*. Oxford: Oxford University Press.

Camp, J. S. & Karmiloff-Smith, A. (in press) *Williams Syndrome.* International Encyclopaedia of Social and Behavioural Sciences.

Acknowledgements

I am very grateful to everyone who helped me along the road to completing my thesis, as well as being indebted to you all for useful discussions and suggestions. Thank you first to my supervisor and mentor Emily Farran, for your enthusiasm and encouragement; patience, understanding and cheerfulness; and knowing when I needed 'reining in'. Your guidance and support has been invaluable. Thank you to my co-supervisors Annette Karmiloff-Smith and Michael Thomas, for their wisdom, insights and advice. A huge thank you is due to the participants, parents and schools involved in the studies, to the Williams Syndrome Foundation and to the Down Syndrome Association, without whom the research could not have been conducted. I am grateful to the Bloomsbury Colleges Symposium for funding the studentship, and to the Institute of Education and Birkbeck College for additional funding. Many thanks to Liz Pellicano for taking me under her wing, and to Sarah Beck for help with questionnaires. Hannah Broadbent, thank you for being my sounding board. Thanks also to other members of and visitors to the CogDev lab, particularly Harry Purser and Suzy Cooper. Kyra, for listening, and making cakes from felt and sponge: what would I have done without you! Matt, my best friend, protector and ally: thank you for believing in me, for spurring me on, and for making me dinner – for once, I don't have *enough* words… The teachers who pushed me onward and upward should be recognised here: Mr Swan, for calling me 'Jo', and Mr Hamblin, for my introduction to Psychology. Finally, thank you to my parents and grandparents for giving me 'roots and wings', and for unfailingly knowing I could do it.

Table of Contents

TABLE OF FIGURES **9**

TABLE OF TABLES **11**

LIST OF FREQUENTLY USED ABBREVIATIONS **12**

CHAPTER 1: INTRODUCTION **14**

1.1 Rationale and approach 14

1.1.1 What is problem solving? *14*

1.1.2 Why study problem solving in neurodevelopmental disorders? *15*

1.1.3 Neuroconstructivism and neurodevelopmental disorders *16*

1.1.4 Using developmental trajectories *16*

1.2 Literature review 17

1.2.1 Williams syndrome (WS) and Down syndrome (DS) *17*

1.2.2 Problem solving *24*

1.2.3 Executive functioning (EF) *36*

1.2.4 Problem solving outside the laboratory *57*

CHAPTER 2: EXPERIMENTAL MEASURES OF EXECUTIVE FUNCTIONS AND PROBLEM SOLVING **63**

2.1 Introduction 63

2.1.1 Links between executive functioning and TOL performance *63*

2.1.2 The current study *64*

2.2 Method 68

2.2.1 Participants *68*

2.2.2 Ethical considerations *69*

2.2.3 Overall design and procedure *70*

2.2.4 Task-specific outlines *71*

2.3 Results 90

2.3.1 TOL coding *90*

2.3.2 TOL scoring *92*

2.3.3 Analyses and parametric assumptions *94*

2.3.4 Matched group comparisons *95*

2.3.5 Relationships between EF and TOL tasks *106*

2.4 Discussion 118

2.4.1 TOL *119*

2.4.2 Chronological age (CA) *121*

2.4.3 RCPM and TOL score *121*

2.4.4 BPVS *123*

2.4.5 *Planning* ... *124*
2.4.6 *Working memory* ... *127*
2.4.7 *Inhibition* ... *131*
2.4.8 *Shifting* ... *138*
2.4.9 *Chapter summary* ... *140*

CHAPTER 3: EXPLORING PERFORMANCE ON THE TOWER OF LONDON TASK ... 142

3.1 INTRODUCTION ... 142
3.1.1 *Problem types* ... *142*
3.1.2 *Scoring* ... *143*
3.1.3 *Aims* ... *148*
3.2 METHOD ... 149
3.3 RESULTS ... 149
3.3.1 *TOL scores by problem type* ... *149*
3.3.2 *EF correlations by problem type* ... *151*
3.3.3 *Timing measures* ... *153*
3.3.4 *TOL behaviours* ... *156*
3.3.5 *TOL correlation analysis* ... *160*
3.4 DISCUSSION ... 163
3.4.1 *Problem type* ... *163*
3.4.2 *Error types* ... *167*
3.4.3 *Timing* ... *168*
3.4.4 *Rule violations (RVs)* ... *173*
3.4.5 *Further measures* ... *174*
3.4.6 *Remaining correlations between measures* ... *176*
3.4.7 *Correlations with CA* ... *177*
3.4.8 *Chapter summary* ... *178*

CHAPTER 4: EVERYDAY MEASURES OF EXECUTIVE FUNCTIONS AND PROBLEM SOLVING ... 180

4.1 INTRODUCTION ... 180
4.1.1 *Problem solving inside and outside the laboratory* ... *180*
4.1.2 *Everyday life in WS and DS* ... *182*
4.1.3 *Measuring everyday life skills* ... *186*
4.1.4 *Behavioral Rating Inventory of Executive Function (BRIEF)* ... *187*
4.1.5 *Problem-Solving Questionnaire (PSQ)* ... *189*
4.2 METHOD ... 192
4.2.1 *Participants* ... *192*
4.2.2 *Materials* ... *195*
4.2.3 *Procedure* ... *196*

4.2.4 Design ... *196*
4.3 Results ... 199
4.3.1 Data handling ... *200*
4.3.2 Percentage score analysis ... *205*
4.3.3 BRIEF analysis ... *212*
4.3.4 PSQ analysis ... *219*
4.4 Discussion ... 235
4.4.1 Group performance across questionnaires ... *235*
4.4.2 Performance on the BRIEF: indices and scales ... *239*
4.4.3 Performance on the PSQ ... *241*
4.4.4 Aspects associated with reaching the solution ... *248*
4.4.5 Exploring and accounting for better DS than WS performance ... *250*
4.4.6 Chapter summary ... *253*

CHAPTER 5: RELATIONSHIPS BETWEEN EXPERIMENTAL AND QUESTIONNAIRE MEASURES ... 254

5.1 Introduction ... 254
5.2 Method ... 254
5.3 Results ... 255
5.4 Discussion ... 260

CHAPTER 6: GENERAL DISCUSSION ... 266

6.1 Aims and overview ... 266
6.2 Discussion of findings ... 267
6.2.1 Summary of findings by chapter ... *267*
6.2.2 Patterns of findings throughout this thesis ... *269*
6.2.3 Considerations from a neuroconstructivist perspective ... *271*
6.3 Limitations and related considerations ... 274
6.3.1 Experimental work ... *274*
6.3.2 Questionnaires ... *280*
6.4 Overall implications and further work ... 282

REFERENCE LIST ... 290

APPENDICES ... 305
Appendix A: Tower of London problem set ... 306
Appendix B: Developmental trajectories with significant relationships in one group only ... 307
Appendix C: BRIEF Example items ... 310
Appendix D: Problem-Solving Questionnaire ... 311

Table of Figures

FIGURE 1.1: SCHEMATIC EXAMPLES OF DEVELOPMENTAL TRAJECTORY PATTERNS BETWEEN GROUPS 17
FIGURE 1.2: THE TOWER OF HANOI (TOH): EXAMPLE OF A 3-DISC PROBLEM 27
FIGURE 1.3: THE TOWER OF LONDON (TOL): EXAMPLE OF A 4-MOVE PROBLEM 28
FIGURE 2.1: SCHEMATIC REPRESENTATION OF TOWER OF LONDON (TOL) APPARATUS 72
FIGURE 2.2: SCHEMATIC REPRESENTATION OF INHIBITION TASK. 75
FIGURE 2.3: PLANNING TASK APPARATUS 80
FIGURE 2.4: APPARATUS FOR PLANNING TASK: HOUSE WITH BALLOON, CAKE AND MILK BOX 81
FIGURE 2.5: GAME BOARDS FOR PHASES TWO AND THREE OF THE PLANNING TASK 82
FIGURE 2.6: INITIAL PRACTICE TRIALS IN PHASE 1 FOR SHAPE AND COLOUR CRITERIA OF THE SHIFTING TASK 84
FIGURE 2.7: EXPERIMENTAL STIMULI IN PHASES 1 AND 2 OF THE SHIFTING TASK 85
FIGURE 2.8: EXAMPLE STIMULI FROM PHASE 3 OF THE SHIFTING TASK: TRAINING AND EXPERIMENTAL 86
FIGURE 2.9: VISUOSPATIAL MEMORY TASK APPARATUS 88
FIGURE 2.10: FREQUENCY DISTRIBUTION OF TOL SCORE ACROSS MATCHED GROUPS 94
FIGURE 2.11: MEAN *(S.E.)* OF TOL SCORE ACROSS MATCHED GROUPS. 96
FIGURE 2.12: TRIALS CORRECT ON THE PLANNING TASK BY GROUP 99
FIGURE 2.13: NUMBER OF SHIFTS MADE ON THE SHIFTING TASK BY GROUP 100
FIGURE 2.14: PERCENTAGE ACCURACY TO SAME AND OPPOSITE CONDITIONS OF INHIBITION TASK BY GROUP 100
FIGURE 2.15: REACTION TIME TO SAME AND OPPOSITE CONDITIONS OF INHIBITION TASK BY GROUP 102
FIGURE 2.16: SPAN MEASURES ACROSS FOUR WORKING MEMORY TASKS BY GROUP 104
FIGURE 2.17: DEVELOPMENTAL TRAJECTORY OF TOL SCORE BASED ON BPVS SCORE 112
FIGURE 2.18: DEVELOPMENTAL TRAJECTORY OF TOL SCORE BASED ON PLANNING TASK SCORE 113
FIGURE 2.19: DEVELOPMENTAL TRAJECTORY OF TOL SCORE BASED ON PROPORTIONAL ACCURACY CHANGE ON THE INHIBITION TASK 114
FIGURE 2.20: DEVELOPMENTAL TRAJECTORY OF TOL SCORE BASED ON RT TO SAME CONDITION OF INHIBITION TASK 115
FIGURE 2.21: DEVELOPMENTAL TRAJECTORY OF TOL SCORE BASED ON RT TO OPPOSITE CONDITION OF INHIBITION TASK 116
FIGURE 2.22: SCATTER PLOT OF BACKWARDS BLOCK SPAN AND TOL SCORE 117
FIGURE 2.23: SCATTER PLOT OF BACKWARDS DIGIT SPAN AND TOL SCORE 117
FIGURE 3.1: GROUP MEANS *(S.E.)* TOTAL TOL SCORE BY PROBLEM TYPE 150
FIGURE 3.2: GROUP MEAN *(S.E.)* OF PLANNING TIME (SECONDS) BY PROBLEM TYPE 154
FIGURE 3.3: GROUP MEANS *(S.E.)* OF TIME PER MOVE (SECONDS) BY PROBLEM TYPE 155
FIGURE 3.4: GROUP MEANS *(S.E.)* OF RULE VIOLATIONS PER ATTEMPTED TRIAL BY ACCURACY 157
FIGURE 3.5: GROUP MEANS *(S.E.)* OF VERBALISATIONS PER ATTEMPTED TRIAL BY ACCURACY 158
FIGURE 3.6: GROUP MEANS *(S.E.)* OF ERROR TYPES 160
FIGURE 4.1: PERCENTAGE OF MISSING ITEMS BY BRIEF SCALE AND GROUP 201
FIGURE 4.2: PERCENTAGE OF MISSING RESPONSES BY QUESTION ON THE PSQ 204
FIGURE 4.3: OVERALL PERCENTAGE SCORES ON THE BRIEF AND PSQ BY GROUP 206

Figure 4.4: Developmental trajectories of BRIEF percentage based on CA 208
Figure 4.5: Developmental trajectories of PSQ percentage based on CA 208
Figure 4.6: Developmental trajectories of BRIEF percentage based on BPVS score 209
Figure 4.7: Developmental trajectories of PSQ percentage based on BPVS score 210
Figure 4.8: Developmental trajectories of BRIEF percentage based on RCPM score 211
Figure 4.9: Developmental trajectories of PSQ percentage based on RCPM score 212
Figure 4.10: Mean item score on the BRIEF scales by group 213
Figure 4.11: Canonical discriminant function scatterplot. 216
Figure 4.12: Group means *(S.E.)* of scores on three BRIEF indices by group 218
Figure 4.13: Group means *(S.E)* of reaching-the-solution score by question and group 219
Figure 4.14: Group means *(S.E)* of shared aspect scores for approaching a problem by question 222
Figure 4.15: Means *(S.E)* of approach aspect scores by question, collapsed across group 224
Figure 4.16: Group means *(S.E)* of response aspect scores 226
Figure 4.17: Means *(S.E)* of response aspect scores by question, collapsed across group.. 228

Table of Tables

TABLE 2.1: PARTICIPANT DEMOGRAPHICS (EXPERIMENTAL WORK)....................69
TABLE 2.2: DESCRIPTIONS OF TOL CODING MEASURES AND BEHAVIOURS 92
TABLE 2.3: EXAMPLE OF TOL SCORING 93
TABLE 2.4: MEAN (SD) OF TRIALS CORRECT SCORE AND ADDITIONAL MOVES ON CORRECT TRIALS....................97
TABLE 2.5: CORRELATIONS BETWEEN PREDICTOR MEASURES AND TOL SCORE FOR EACH GROUP 108
TABLE 3.1: CORRELATIONS OF EF MEASURES WITH TOL SCORE FOR EASY, MEDIUM AND HARD PROBLEMS.. 152
TABLE 3.2: CORRELATIONS BETWEEN BEHAVIOURAL TOL MEASURES FOR THE TD GROUP.. 161
TABLE 3.3: CORRELATIONS BETWEEN BEHAVIOURAL TOL MEASURES FOR THE DS GROUP....................162
TABLE 3.4: CORRELATIONS BETWEEN BEHAVIOURAL TOL MEASURES FOR THE WS GROUP....................163
TABLE 4.1: QUESTIONNAIRE PARTICIPANT DEMOGRAPHICS 193
TABLE 4.2: DEMOGRAPHICS OF SUBSETS OF QUESTIONNAIRE PARTICIPANTS WITH MENTAL AGE DATA 194
TABLE 4.3: PSQ QUESTION TOPICS....................197
TABLE 4.4: ASPECTS OF PROBLEM SOLVING SHARED ACROSS QUESTIONS IN THE PSQ 198
TABLE 4.5: FACTOR LOADINGS OF BRIEF SCALES FOR DISCRIMINANT FUNCTIONS ANALYSIS 216
TABLE 4.6: CORRELATIONS BETWEEN PSQ APPROACH ASPECTS AND REACHING THE SOLUTION BY GROUP....................230
TABLE 4.7: CORRELATIONS BETWEEN PSQ RESPONSE ASPECTS AND REACHING THE SOLUTION BY GROUP....................231
TABLE 4.8: CORRELATIONS BETWEEN BRIEF SCALE SCORES AND REACHING-THE-SOLUTION SCORE ON THE PSQ FOR ROUTINE AND NOVEL PROBLEMS.. 233
TABLE 4.9: CORRELATIONS BETWEEN TASK-MONITOR AND SELF-MONITOR SCALES AND REACHING-THE-SOLUTION SCORE ON THE PSQ FOR ROUTINE AND NOVEL PROBLEMS.. 234
TABLE 5.1: CORRELATIONS BETWEEN MA/EF MEASURES AND ROUTINE AND NOVEL REACH-THE-SOLUTION SCORE ON THE PSQ....................258
TABLE 5.2: CORRELATIONS BETWEEN EF AND QUESTIONNAIRE MEASURES FOR TD, DS AND WS GROUPS. 259

List of Frequently Used Abbreviations

BRIEF[1]	'Behavior Rating Inventory of Executive Function'
CA	Chronological Age
DS	Down Syndrome
EF	Executive Function(s)/Functioning
MA	Mental Age
PSQ	Problem-Solving Questionnaire
RT	Reaction Time
TD	Typically Developing/Typical Development
TOH	Tower of Hanoi
TOL	Tower of London
WS	Williams Syndrome

[1] The BRIEF originated in the USA and as such, when referring to the name of the BRIEF and its subscales, the relevant US spellings are adopted accordingly.

“All life is problem solving” (Karl Popper, 1999)

CHAPTER 1: INTRODUCTION

The overall aim of this thesis is to investigate problem-solving abilities in typical and atypical development. The first section of this chapter will address the rationale and theoretical approach taken. In the second section, relevant literature will be reviewed in order to: introduce the atypical populations of interest and the topic of problem solving; consider the processes that might underlie problem solving and its development; and discuss relevant issues that arise in the measurement of such constructs. The chapter will conclude with a statement of the specific aims of the research and an outline of the structure and content of this thesis.

1.1 Rationale and approach

1.1.1 What is problem solving?

While problem-solving research can be found in the literature review section of this chapter, an introductory note is included here to clarify the concept for the reading of this thesis. The conceptualisation of problem solving is not straightforward. In general usage it implies that there is an out-of-the-ordinary aversive situation that requires some kind of action to be resolved. For example, one may have a moral dilemma, or a broken-down car engine. In academic usage the interpretation tends to be broader, so that every task which we undertake is seen as the solving of a problem, for example, how do I attain the goal of providing an evening meal? (Anderson, 1993; Davidson & Sternberg, 2003). As Anderson (1993) points out, rather than only including activities that feel effortful in the category of problem solving, for those who study it, “all higher level cognition is problem solving” (p. 39). Conversely, in experimental paradigms, the novelty of a posed problem is often emphasised as important, in order to facilitate the study of the generation of new solutions. In these problem-solving tasks, the onus is on the solver to produce the solution (e.g., Klahr & Robinson, 1981; Shallice, 1982). In this thesis, a problem is treated as a situation in which there is a goal to be reached, and in which the solver must find the way to reach it.

As such, the central experimental problem-solving task employed in this thesis is the Tower of London (TOL; Shallice, 1982). The TOL is often referred to as a problem-solving task (e.g., Berg & Byrd, 2002), but is also included in several test batteries assessing executive functioning (EF; e.g., the Cambridge Neuropsychological Test Automated battery, CANTAB; Owen, Downes, Sahakian, Polkey, & Robbins, 1990) and in studies aiming to investigate complex EF skills (Huizinga, Dolan, & van der Molen, 2006). Thus, the TOL could be argued to reflect either problem solving or complex EF abilities. As EF is thought to underlie goal-directed behaviour, the difference between the use of these two terms is subtle, if present at all. Here, the TOL is seen as reflecting problem solving because the participant must produce the solution themselves (cf. Klahr & Robinson, 1981), rather than only following rules for responding to stimuli, as might be required in an EF task measuring, for example, inhibition. Thus, in this thesis, the TOL is conceptualised as a problem-solving task, which is very likely to rely on a collection of related but distinct, lower-level EF processes (inhibition, working memory; citations in Section 1.2.3).

1.1.2 Why study problem solving in neurodevelopmental disorders?

Problem solving is a ubiquitous and crucial part of everyday life, encountered in a wide variety of contexts and encompassing a wide range of capacities. Problem solving has been well documented in typical adults (e.g., Davidson & Sternberg, 2003; Robertson, 2001; Simon, 1975; Welsh, Satterlee-Cartmell, & Stine, 1999) and children (e.g., Bull, Espy, & Senn, 2004; Fagot & Gauvain, 1997; Klahr, 1985; Klahr & Robinson, 1981; Welsh, 1991; Winsler & Naglieri, 2003). However, in individuals with neurodevelopmental disorders, while a considerable amount of work has addressed specific cognitive abilities (see, for example, Mervis & John, 2010; Vicari, 2006 for reviews) and some studies have used problem-solving tasks as measurements of planning within a wider assessment (e.g., Menghini, Addona, Costanzo, & Vicari, 2010), there has been a limited amount of research addressing the ability of individuals to bring specific skills together in order to solve a problem. The aim of research with these groups must ultimately be to support the daily lives of individuals with the disorders. Therefore, it is imperative to develop our understanding, not only of isolated abilities, but how those abilities are used: not only in well-controlled experimental settings, but in the real world, where individuals with developmental disorders are attempting to solve problems every day. In this thesis, these are the topics that are addressed.

1.1.3 Neuroconstructivism and neurodevelopmental disorders

In this thesis the study of genetic disorders is approached from a neuroconstructivist perspective. Both Williams syndrome (WS) and Down syndrome (DS) are known to be of genetic aetiology (see below). The nativist approach to development would lead to the conceptualisation of the resulting atypical phenotype as the direct product of some missing components (or 'modules'; Fodor, 1983) in a set of abilities that rely solely (or mainly) on the disrupted genetic material (see critical discussion in Karmiloff-Smith, 1998). In contrast, neuroconstructivism views development as dynamic and interactive, with low-level impairments to the young and poorly specialised system making themselves felt to different degrees in different domains as the system matures, learns and becomes more specialised through 'progressive modularisation' (Karmiloff-Smith, 1998). One tenet of this approach is that similar-looking behaviours can be attained though different cognitive processes (e.g., in the domain of face processing; Karmiloff-Smith et al., 2004). Another is that researchers should take a developmental approach, examining outcomes across development in order to understand how phenotypes are reached, and comparing outcomes across neurodevelopmental disorders (Karmiloff-Smith, 2009). Both of these methodologies are used in this thesis.

1.1.4 Using developmental trajectories

The developmental trajectory approach is one of the central statistical techniques employed in this thesis. Stemming from the requirement to take a developmental view, it allows task performance to be plotted, for each group, against another measure, to ascertain a developmental pathway which can then be compared across groups. Here, performance on an outcome task (the TOL) is analysed with respect to various predictor measures (chronological age, CA; mental age, MA;

executive function (EF) task performance). A trajectory is constructed for each group of participants, and compared using a technique derived from analysis of covariance (ANCOVA). Thomas et al. (2009) presented guidance for its use. Comparing trajectories allows the determination of whether the scores on the outcome measure differ across groups at a particular point on the trajectory, as well as whether the gradients of the lines differ; that is, whether the rate of development of the outcome with respect to the predictor is similar to the typical group or looks atypical. Some examples of different types of trajectory (Thomas et al., 2009) can be seen in Figure 1.1.

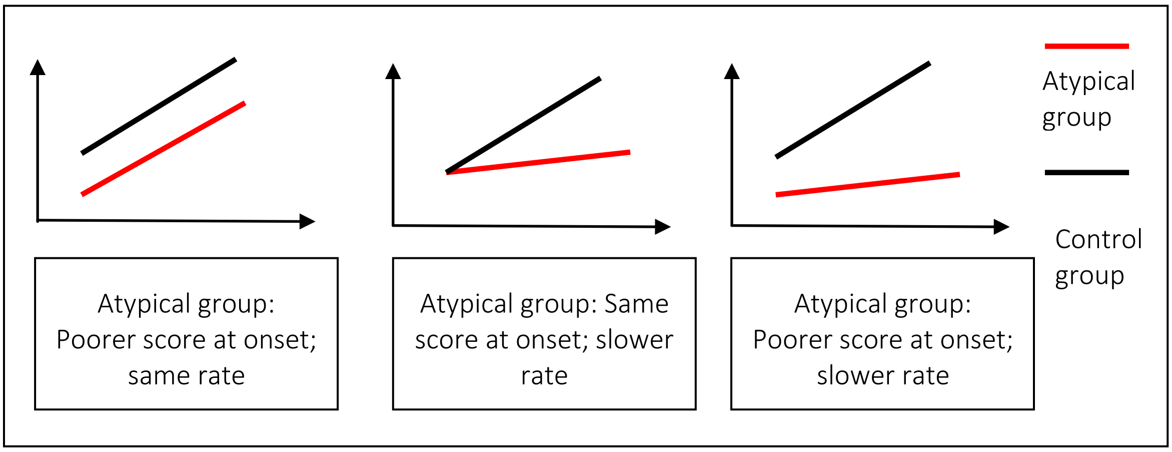

Figure 1.1: Schematic examples of developmental trajectory patterns between groups

In the next section of the chapter, the two neurodevelopmental disorders to be studied will be introduced, before reviewing research into problem solving (studies specifically addressing EF, and TOL problem solving, in WS and DS are discussed later in the chapter, after these topics have been introduced).

1.2 Literature review

1.2.1 Williams syndrome (WS) and Down syndrome (DS)

1.2.1.1 Prevalence and genetics of WS and DS

WS is a rare developmental disorder, with a prevalence of 1 in 20, 000 live births (Morris & Mervis, 1999; Pober, 2010), although a recent estimate is considerably higher, at 1 in 7500 (Strømme, Bjømstad, & Ramstad, 2002). Between 26 and 28 genes are deleted from chromosome 7q11.23 (Pober, 2010), including the *ELN* gene for elastin (Ewart et al., 1993), which contributes to some of the medical problems seen in WS (Pober, 2010). While there is much interest in the contribution of genetic deletions to the cognitive phenotype in WS, the relationship between genotype and phenotype is a complex and multidimensional one (Gray, Karmiloff-Smith, Funnell, & Tassabehji, 2006; Karmiloff-Smith et al., 2012) due to intricate interactions at different levels of description, eventually leading to an individual's dynamic phenotype (Karmiloff-Smith, 1998; 2006).

By contrast, DS, initially described by Down (1866), is a relatively common chromosomal genetic disorder occurring in 1 in 650-1000 births (Bittles, Bower, Hussain, & Glasson, 2007). The most common form, trisomy 21, involves a *de novo* occurrence of an extra copy of the entire chromosome 21 (Cody & Kamphaus, 1999). LeJeune, Gautier and Turpin (1959) discovered this extra chromosome. Approximately 94% of DS instances occur in this way, with the remaining 6% being familial (Pennington, Moon, Edgin, Stedron, & Nadel, 2003). Attempts to link the DS phenotype with specific genes are underway, e.g., using mouse models (Crnic & Pennington, 2000), although this process is dependent on knowledge of which genes, being overexpressed, would affect brain development and which would be expressed at the appropriate point in ontology to affect development (Pennington et al., 2003).

1.2.1.2 Cognitive phenotype in WS and DS

In this section, the cognitive phenotype of individuals with WS and DS is reviewed, both to provide a broad introduction to the syndromes and in recognition that problem solving is likely to draw on many different abilities.

People with WS have a distinctive facial appearance and mild to moderate learning difficulties, alongside several medical irregularities (e.g., Bellugi, Wang, & Jernigan, 1994; Donnai & Karmiloff-Smith, 2000; Jones & Smith, 1975). A friendly nature and good grasp of language were initially observed (von Arnim & Engel, 1964) while more recent studies find individuals with WS to be 'hypersocial' (Jones et al., 2000), although they often experience anxiety (Stinton & Howlin, 2012), isolation (Jawaid et al., 2011) and have problems with peer relationships and social skills (Campos & Sotillo, 2012; Stinton & Howlin, 2012; Stojanovik, 2006).

The cognitive profile in WS is uneven (Mervis et al., 2000), with generally better verbal than visuospatial abilities, although this unevenness has been emphasised to different degrees by different researchers. Some report normal face processing (Tager-Flusberg, Plesa-Skwerer, Faja, & Joseph, 2003) and language skills, against a backdrop of severely impaired visuospatial construction (Bellugi, Lichtenberger, Jones, Lai, & St. George, 2000) while others observe relative strengths and weaknesses, both between and within domains. For example, difficulties with pragmatics, grammar and other areas of language have been found (Brock, 2007; Martens, Wilson, & Reutens, 2008) while there is evidence for atypical neurocognitive face processing in WS which nonetheless still produces behavioural scores in the normal range (Karmiloff-Smith et al., 2004). Also, concrete receptive vocabulary is a relative strength in children with WS (Mervis & John, 2008) while difficulties with spatial language have been identified (Laing & Jarrold, 2007; Phillips, Jarrold, Baddeley, Grant, & Karmiloff-Smith, 2004). In children, relational vocabulary was found to be comparable to visuospatial construction abilities (Mervis & John, 2008) which is an area of extreme weakness for this population, for example on block design tasks (e.g., Farran, Jarrold, & Gathercole, 2001; Hoffman, Landau, & Pagani, 2003). These abilities are poor, even within the already weak domain of nonverbal skills (Farran & Jarrold, 2003). Better verbal than visuospatial short-term memory (STM) is generally found for this group, in contrast to DS (e.g., Jarrold, Baddeley, & Hewes, 1999): see below.

Generally, individuals with WS score between 40 and 90 on the Wechsler Intelligence Scale for Children-Revised, WISC-R (Semel & Rosner, 2003) with full-scale IQ remaining stable throughout adulthood (Howlin, Elison, Udwin, & Stinton, 2010). However, the uneven WS profile means that full-scale IQ scores can be misleading (see Farran & Jarrold, 2003 for a discussion of methodological issues in WS testing). When studies measure abilities separately, the uneven profile becomes apparent, e.g., Mervis and John (2010) found particular weaknesses on spatial IQ scores on the Differential Ability Scales – 2nd edition (DAS-II; Elliott, 2007). In addition, Mervis and John (2010) point out the challenges of interpreting the results of specific IQ tests, which differ in how the samples are normed. A test which does not provide low enough norms for standardisation will not be able to capture the performance of individuals with WS; the Wechsler IQ tests are included as a case in point. Thus, caution should be used when interpreting IQ scores in WS. These uneven levels of proficiency in WS are important to take into account when considering how problem solving might be approached: problem solving involves the combination of many different skills, so some may be relied upon more than others.

Most children with DS have mild, moderate, severe or profound learning difficulties, which impacts on cognitive development (Cody & Kamphaus, 1999). IQ (i.e., performance in relation to developmentally-appropriate expectations) declines over developmental time in childhood, remaining relatively stable in adulthood (Carr, 2012). Generally adults with DS have with a low IQ (25-55) (Gibson, 1978) although ability levels are by no means homogenous (Cody & Kamphaus, 1999). Tsao and Kindelberger (2009) also emphasise the heterogeneity of ability levels in children with DS.

Individuals with DS are known to be friendly and to show a liking for social stimuli and interaction, much as people with WS do (Porter, Coltheart, & Langdon, 2007). Children with DS have early social strengths and are able to maintain friendships in childhood but socio-cognitive tasks are more difficult, possibly because of EF demands (see Fidler & Nadel, 2007 for a review). The DS cognitive profile is uneven and complex, and is reviewed by Vicari (2006). In infancy, language learning is impaired, with slow vocabulary acquisition compared to CA, but at the same level as MA-matched controls and as infants with WS (Paterson, Brown, Gsodl, Johnson, & Karmiloff-Smith, 1999). Later in life, language is generally weak, although comprehension skills are stronger than production skills (Vicari, 2006). Children with DS show poorer grammatical skills than those with WS although both groups demonstrate delayed language overall on a parental report measure of language development (Singer Harris, Bellugi, Bates, Jones, & Rossen, 1997). However, compared to their poor language abilities, visuospatial abilities are a relative strength in DS (Vicari, 2006).

Numerous studies have found evidence of working memory deficits in the verbal domain, with stronger visuospatial short-term memory (STM) skills in DS (e.g. Laws, 2002) while the opposite pattern is seen in WS (Jarrold et al., 1999; Wang & Bellugi, 1994). Backwards span tasks, purported to recruit the central executive, are also poor in DS compared to TD controls and other intellectually impaired groups (Vicari, Carlesimo, & Caltagirone, 1995) and there is evidence for impaired explicit memory against a background of stronger implicit memory (Vicari, Bellucci, & Carlesimo, 2000). Reaching strategies on an object retrieval problem-solving task were found to be less efficient in a group of toddlers with DS than a CA- and MA-matched group with developmental disabilities and a MA-matched typical control group, indicating early problem-solving difficulties in DS (Fidler, Philofsky, Hepburn, & Rogers, 2005).

Auditory processing in children with DS is slower than that of both CA- and MA-matched controls (Lincoln, Courchesne, Kilman, & Galambos, 1985), while information processing is slow in DS overall (Gibson, 1991, cited in Cody & Kamphaus, 1999). Difficulties with approaching the learning of new skills and holding on to existing ones have been identified in children with DS (Wishart, 1993a). Performance on verbal and spatial long-term memory (LTM) tasks is impaired in adolescents with DS compared to MA-matched controls (Pennington et al., 2003), as well as in infants compared to CA controls (Mangan, 1992) and adults (e.g., Devenny et al., 1992, both cited in Nadel, 2003) which is in line with findings of hippocampal abnormality in DS (Nadel, 2003).

1.2.1.3 Brain structure and function in WS and DS

Both WS and DS brains show an overall reduction in size compared to CA-matched controls (Reiss et al., 2000; Weis, 1991). Children with DS (aged 2-8 years) have smaller brain volumes compared to children with developmental language delay and Fragile X syndrome as well as to controls (Kates, Folley, Lanham, Capone, & Kaufmann, 2002). Atypicalities of brain structure and function are seen for both syndromes (see, e.g., Jackowski et al., 2009; Nadel, 2003 for reviews).

Evidence for frontal (Owen et al., 1990; Stuss & Benson, 1984) and prefrontal cortex (PFC) (Ball et al., 2011; Koechlin, Ody, & Kouneiher, 2003; Mushiake et al.) involvement in goal-directed behaviour (i.e., problem solving) in typically developing individuals is ongoing (although see Rowe, Owen, Johnsrude, & Passingham, 2001 and section 1.2.3.5). Indeed, cerebral blood flow increased in the left PFC during a TOL task (Morris, Ahmed, Syed, & Toone, 1993). Various atypicalities of these brain areas have been found in the two syndromes of interest. WS brains have more complex cortical folds than is normal (Gaser et al., 2006; Schmitt et al., 2002), and Fahim (2012) found increased cortical complexity in frontal lobes in children with WS, amongst other areas. Chiang et al. (2007) noted that the prefrontal and orbitofrontal areas of WS brains, as well as several other areas including the amygdala and cerebellum, were not substantially reduced in size. However, grey matter in orbitofrontal regions has been reported to be both increased and decreased, depending on the technique used in analysis, which stems from the problems in comparing WS and typical brains, which are different shapes (Eckert et al., 2006). In DS, the frontal lobes show decreased growth compared to typical controls in infancy and the PFC is identified as one of the main areas of neuropathology in DS, amongst other areas (Fidler & Nadel, 2007; Nadel, 2003). There is also decreased gyrification in the frontal lobe in adults with DS (Lögdbergi & Brun, 1993). Unterrainer and Owen (2006) noted, importantly, that the frontal areas of the brain are connected to several other areas, and function using those connections, so goal-directed behaviour is unlikely to depend only on frontal regions. For example, unusual patterns in the way in which the amygdala and PFC converse are thought to underlie the atypical social behaviours seen in WS (i.e., fearful reactions to non-social stimuli but not to threatening social stimuli) (Meyer-Lindenberg, Hariri, et al., 2005; but see Porter et al., 2007). Other brain areas also show abnormalities in WS and DS, as reviewed below.

In children with WS, Fahim et al. (2012) also found increased cortical complexity and gyrification index in the parietal lobe, alongside decreased cortical volumes and surface area, compared to CA-matched typical controls. Frontal-parietal areas are thought to be involved in the visual control of movement, and also show atypicalities in WS (Hocking, Bradshaw, & Rinehart, 2008 and references therein). Indeed, there are well-recognised atypicalities of the parietal lobe in WS, thought to be implicated in visuospatial functioning (Chiang et al., 2007; Eckert et al., 2006; Weis, 1991). Reiss and colleagues have also found atypicalities in areas thought to be implicated in visual processing in WS: reduced grey matter volume of the occipital cortex and thalamus (Reiss et al., 2004) and occipital lobe asymmetries (Reiss et al., 2000). Chiang et al. (2007) demonstrated reduced volumes of some brain areas, including occipital and parietal regions as well as the basal ganglia. Connectivity is unusual in the WS brain: notably, reduced flow of information to dorsal-stream regions has been found during visuospatial construction tasks (Meyer-Lindenberg et al., 2004) which are known to be particularly poor in WS (see above). Other areas that are often found to show atypicalities are the hippocampus (involved in spatial cognition amongst other functions: see below) and corpus callosum, which connects the hemispheres and is thought to play a role in problem solving amongst other skills (see Lögdbergi & Brun, 1993 for a review). Small corpus callosum areas and atypical hippocampal shapes were noted in WS, compared to typical age-matched controls' brains (Meyer-Lindenberg, Mervis, et al., 2005; Schmitt, Eliez, Warsofsky, Bellugi, & Reiss, 2001).

At birth, DS brains are relatively comparable to typical brains, but during infancy several abnormalities become evident, including a reduced overall brain size, disproportionately smaller brain stem and cerebellum, and delayed myelination in some cases (as well as the decreased frontal lobe growth mentioned above) (for a review, see Fidler & Nadel, 2007; Nadel, 2003). This emergence of abnormality does not always reflect the characteristics of the group as a whole, but rather, the number of individuals within the group who show atypicalities, highlighting the heterogeneity of the disorder (Fidler & Nadel, 2007). Hippocampal volumes have been found to be reduced in MRI studies (Jernigan, Bellugi, Sowell, Doherty, & Hesselink, 1993; Pinter et al., 2001). Above age 35, the risk of early onset dementia is higher than in the typical population (Bush & Beail, 2004). All adults with trisomy 21 present with the brain pathology that is seen in Alzheimer's disease, while only around half go on to develop the associated dementia symptoms (Nadel, 2003). In addition to the PFC, notable areas of neuropathology in DS are the hippocampus, implicated in spatial cognition, memory consolidation and flexible learning, and the cerebellum, with a less clearly defined role which potentially includes motor skills and learning conditioned responses (Fidler & Nadel, 2007; Nadel, 2003). Motor development is slow compared to controls, but milestones are reached in the same order (Vicari, 2006).

In the next section, the topic of problem solving will be introduced, and the processes underlying it will be explored. For an overview of executive functioning abilities in WS and DS, see Section 1.2.3.5.

1.2.2 Problem solving

1.2.2.1 Introduction to problem solving

Research into problem solving stretches back to the work of the Gestalt psychologists and beyond, when problems that were posed usually required a 'eureka' moment, or insight, into how the problem should be solved (Dunbar, 1998). An example is the famous pendulum problem (Maier, 1931, cited in Eysenck & Keane, 2000) in which a participant is asked to tie together two strings hanging from the ceiling which are too far apart to be held at the same time: a process known as 'restructuring' must occur (prompted by the examiner 'accidentally' moving a string as they walk past) to enable the participant to produce the solution through changing the way in which the problem was represented (Eysenck & Keane, 2000) and tie one of the objects provided to one string so it can be swung towards the other and caught. Goal-directed behaviour has been observed in animal studies, e.g., in crows (Clayton, 2007) as well as in infants (e.g., Chen, Sanchez, & Campbell, 1997) and related areas, such as tool innovation in children (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011). Other avenues of research include presenting analogous problems known as isomorphs and investigating the relations between them, and studying the problem solving of experts, such as chess players (Kahney, 1986b). Some ways of capturing processes during problem solving include verbal protocols ('thinking out loud') (e.g., Newell & Simon, 1972) and analysing the gestures people produce (Garber & Goldin-Meadow, 2002). As mentioned above, goal-directed behaviour is thought to be mediated by the PFC.

1.2.2.2 The information-processing approach

A prominent approach to studying problem solving is known as information processing. Perhaps its best-known proponents, Newell and Simon (1972)'s seminal theory of problem solving uses 'production systems', a cognitive science approach to both capture human cognition and develop intelligent machines (Schunn & Klahr, 1998). This is based on the existence of a set of condition/action (IF/THEN) rules which are stored, in humans, in long-term memory, and interact with conditions in the environment, stored in 'working memory' to produce behavioural responses. A simple example is that the current condition (e.g., it is raining) is stored in the 'working memory' and compared to the production rules for a match. When a match is found, the appropriate action linked to the existing condition is activated (e.g., put your hood up) (Kahney, 1986a). These processes can be combined in complex ways to model complex task performance and cognition as a whole: for an overview of production systems, see Schunn and Klahr (1998) and Young (2001). Newell and Simon (1972) identified rules (productions) from participants' problem-solving behaviour and verbal protocols. In their problem space theory, the individual who is solving the problem (the solver) moves from a start state to a goal state using a series of operators (moves). It is based on the idea that the journey from the start of a problem to its finish and all the possible moves along the way can be represented abstractly in a problem space (as such it is an example of a symbolic approach to cognition). The term ‘problem space’ refers to the solver’s representation of the problem (termed ‘task environment’ in its objective sense) in which the actual solving will occur (Simon & Newell, 1971). Problem solving is conceptualised as the search of a problem space (Dunbar, 1998). A problem space can be very large when there are many possible combinations and permutations of moves, and it has been suggested that humans are only able to represent limited amounts of it at once (Dunbar, 1998). Human processing is also thought by production systems theorists to be serial (rather than parallel) with a limited short-term memory capacity underlying the process, and easy, fast access to unlimited long-term memory stores that nevertheless take a large amount of time to store information (Simon & Newell, 1971).

Problem solving task performance can be modelled and compared to human data to assess the success of the model. Several problem-solving models have been based on the production systems (symbolic) approach, e.g., Norman and Shallice (1980); the General Problem Solver (GPS; Newell, Shaw, & Simon, 1959) and Adaptive Character of Thought (ACT; Anderson, 1993, 1996), which is a theory of cognition, used to model problem solving (see, e.g., Altmann & Trafton, 2002). More recently, symbolic approaches have been extended to form hybrid symbolic-connectionist computational models (see, e.g., Lebiere & Anderson, 2008). Production Systems theorists often consider problem solving on tasks which are easily measured and characterised; that is, problems with a finite problem space. Two classic examples of such tasks are the Tower of Hanoi (TOH) and the Tower of London (TOL).

1.2.2.2.1 The Tower of Hanoi (TOH)

The TOH problem was developed by Simon (1975). In the task, participants are presented with a number of different-sized discs, stacked on the leftmost of three wooden posts (the start state, shown in Figure 1.2), and their task is to move the collection of discs to the rightmost post (the goal state), while following certain rules, such as only placing smaller discs on top of larger discs and only moving one disc at a time. The operators are the moves that can be applied, while the task is constrained by the rules. As Klahr and Robinson (1981) and Shallice (1982) point out, the challenge is in finding the correct and minimum sequence of moves to reach the goal state. A schematic diagram of a 3-disc TOH puzzle can be seen in Figure 1.2.

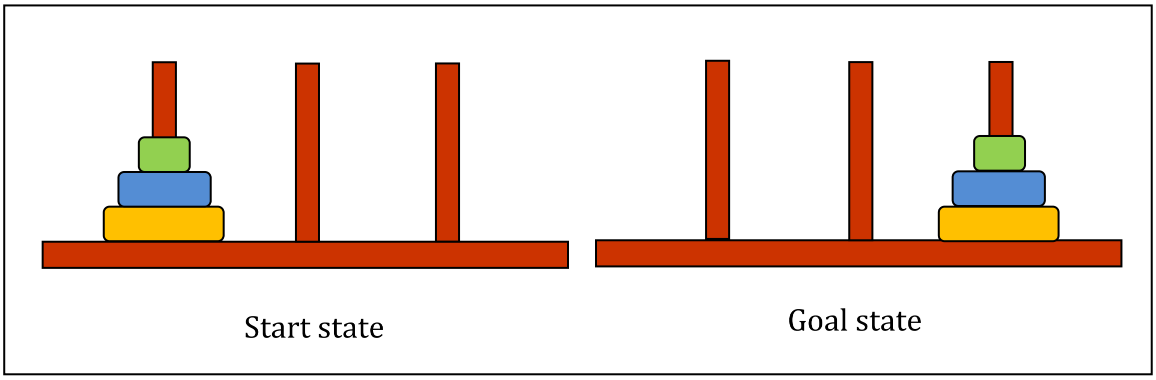

Figure 1.2: The Tower of Hanoi (TOH): example of a 3-disc problem

While task difficulty can be varied using different numbers of discs (e.g., Piaget, 1976) or pegs, there is a particular strategy for solving the TOH puzzle. It is known as a goal recursion strategy, in which temporary smaller stacks of discs must be produced to allow larger discs to move (here, the small disc needs to be stacked on the goal peg temporarily and then the medium and small discs need to be stacked on the middle peg so that the larger one can be moved to the rightmost peg).

1.2.2.2.2 The Tower of London (TOL)

The TOL was developed by Shallice (1982) from the TOH task, so that neuropsychological patients could be tested on a task in which the level of difficulty could be varied; indeed, they can range from very easy to extremely difficult. Three pieces ('beads') of different colours are presented on three pegs of varying height, which can hold three, two or one pieces, which need to be rearranged, one at a time, to match a goal configuration. When Shallice designed the task, difficulty was measured by the minimum number of moves required to reach the solution, with more difficult problems generally requiring more moves to be made between the start state and the goal state. A schematic diagram of a 4-move problem can be seen in Figure 1.3.

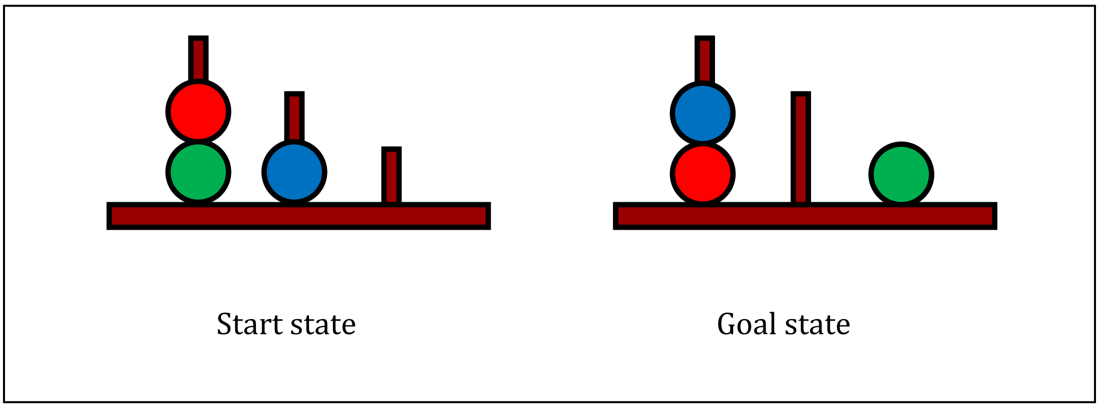

Figure 1.3: The Tower of London (TOL): example of a 4-move problem

The TOL, due to its potential for systematically varying the level of difficulty of individual problems, is thought to be suitable for children and clinical populations. This was one of the main reasons for its selection as the main experimental task for this thesis. It was thought to be relatively free of complex perceptual demands and appealing for participants, with a range of readily observable behaviours and error types which could be expected to provide insight into the problem-solving processes being employed, as well as an individual's ability to correctly solve a problem. Thirty years from Shallice's original publication, measures of TOL difficulty and proficiency now vary widely in research studies. In order to inform the interpretation of TOL research, the next section outlines these variations, before addressing the types of strategies that are thought to be employed, within the information-processing framework.

1.2.2.3 Properties of the TOL task

The properties of the TOL task allow much variation in the specific problems that are administered. While this is a strength of the design, and indeed one of the reasons for its development (Shallice, 1982), the task paradigm and the ways in which the task may vary also means that there are many different characteristics of problems that may vary, rather than a linear increase in difficulty which might be seen in, for example, memory span tasks. The TOL, as it exists now, can differ along several distinct dimensions: the problem set, or the specific arrangements of start and goal positions that participants are asked to make their way between; the form of the task, that is, whether it is broadly similar to Shallice's original design or not; and the way in which it is administered, with respect to variations in instructions, procedure and scoring. Each of these dimensions will be considered in this section.

1.2.2.3.1 Problem set (parameters)

While TOL problems consist of a start and goal state and the requirement to move from one to the other in a series of moves, the most intuitive measure of difficulty is the number of moves that must be made in order to get from the start to the goal. This is known as the minimum number of moves, and was the original measure of difficulty employed by Shallice (1982): the more moves needed, the more difficult the problem. However this is not always adequate (Berg & Byrd, 2002; Berg, Byrd, McNamara, & Case, 2010; Kaller, Unterrainer, Rahm, & Halsband, 2004) and there are numerous other parameters, some of which are reviewed below (for a comprehensive discussion, see Kaller, Rahm, Köstering, & Unterrainer, 2011).

The moves made from start to finish are not equal. Take the example in Figure 1.3 above. The task can be considered in terms of one overarching goal, which can be broken down into three subgoals: that is, when each piece is placed in the particular position in which it needs to be (its 'goal position'), we can say that one subgoal has been reached (it is also possible to break subgoals down further, for example to have a subgoal of removing an obstacle to allow us to move a particular piece, but here we will consider subgoals as piece placements). Thus, placing a piece in its goal position is known as a 'goal move', while any other moves are termed intermediate moves. One TOL parameter which affects performance is 'search depth': the number of intermediate moves that occurs before the goal move, to allow a subgoal to be reached (Spitz, Webster, & Borys, 1982).

Each problem has an optimum path through the problem space, which allows it to be solved in the minimum number of moves. Sometimes there will only be one path, and sometimes there will be more than one. This affects task difficulty (Berg et al., 2010; Kaller et al., 2011). Moreover, sometimes the optimum move is to temporarily move a piece away from its eventual goal position. These are variously known as indirect moves, counterintuitive moves, or goal-subgoal conflict moves (Bull et al., 2004; Kaller et al., 2011) and are referred to here as counterintuitive moves. Preplanning time was found to increase for problems with four counterintuitive moves compared to those with no counterintuitive moves (Phillips, 1999) although their number was confounded with the number of moves to the solution. This is indicative of the way in which the various parameters of the problems are not always independent: for example, as the number of counterintuitive moves increases, so will the overall minimum moves to solution.

The configuration of the goal state can also vary: this can be tower-ending, with all the pieces on top of each other on one peg; flat-ending, with each piece on a different peg, or partial tower-ending, with two pieces on one peg and one piece on another. Flat-ending problems are more difficult to solve than tower-ending ones, because there is no 'ambiguity' in the order in which the pieces must be placed (Klahr & Robinson, 1981). Partial tower-ending problems lie between flat- and tower-ending configurations in terms of difficulty (Spitz et al., 1982).

1.2.2.3.2 Forms of the TOL task

Berg and Byrd (2002) give an overview of the different ways in which researchers have presented the TOL problem, and have classified the different approaches. Some tasks, classified as the 'Shallice TOL' resemble the original in terms of design (for example, three pieces that must be moved on three pegs of decreasing height), with the problem space and rules of the task intact, although they may look slightly different or provide different problem sets. In contrast, Berg and Byrd state that other forms of the TOL change the task demands to such a degree that it is difficult to compare across versions: isomorphs of the problem include a version of the TOL that uses "pockets" or "socks" (Owen et al., 1990) that the pieces are dropped down into, rather than pegs onto which they are threaded. This type of TOL task is used in the CANTAB and is known as the Stockings of Cambridge (SOC) task. Other variants adapt the number of pieces, number of pegs and height of the pegs, which, as this alters the problem space of the task, make it impossible to effectively compare outcomes from isomorphs to outcomes from the Shallice TOL (Berg & Byrd, 2002).

1.2.2.3.3 Procedural variations

An additional important way in which researchers vary in their administration of the TOL is the procedural approach that is taken, such as the instructions given. This is likely to affect the way in which the task is completed (Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003) as well as the EF skills that are relevant. For example, Bull et al. (2004) found that inhibition (see Section 1.2.3) was associated with TOL but not TOH performance in their study, and attribute this to differences in task administration: on the TOL, but not the TOH, children were told the minimum number of moves before each trial. As described below, tower-based tasks can be solved with a period of pre-planning before execution begins, or by just 'jumping in' and solving the puzzle piece-by-piece, known as a perceptual strategy. Bull and colleagues (2004) suggest that information about the number of moves on the TOL would encourage children to inhibit the natural response to just start moving the pieces. The perceptual strategy, however, has also been suggested to invoke inhibition abilities. This is because the tendency to make the move which brings one perceptually closer to the goal state must be inhibited in favour of the better long-term option (e.g., Miyake et al., 2000).

1.2.2.3.4 Scoring considerations

A variety of different scoring systems have also been used in different studies (for a discussion, see Berg & Byrd, 2002), including the number correct; the number correctly solved in the particular minimum number of moves for that problem (known as perfect solutions); the number of extra moves made beyond the minimum, as a measure of efficiency; the number of attempts taken to solve a problem, when repeated attempts are allowed; planning time (time taken before the first move); time taken to solve the remainder of the problem; rule violations; and so on.

Baker, Segalowitz, and Ferlisi (2001) compared TOL scores developmentally that had arisen from using two alternative scoring methods. One was employed by

Krikorian, Bartok, and Gay (1994) and was termed the 'accuracy scoring method': trials are awarded three, two or one point(s) if completed correctly on the first, second or third attempt, respectively. The other was called the 'time-sensitive scoring method' by Baker and colleagues, and is calculated by subtracting the number of attempts from the weighted solution time (used by Anderson, Anderson, & Lajoie, 1996). The outcomes using the two scoring systems were highly correlated for participants aged seven years ($r = 0.86$) but had dropped to 0.47 by adulthood. Baker and colleagues pointed to the importance of the scoring method used for interpretation of outcomes and asserted that the different scoring systems must be measuring different constructs in adults and children. Even when comparing samples of a similar age, using the 'accuracy scoring method' it seems that group differences could be produced either by differences in the number of trials that were solved correctly within three attempts, or in the number of attempts needed to solve a trial. It is clear how differences in scoring procedures serve to complicate the literature on the TOL task, which highlights the importance of using explicit scoring measures for a task as complex and widely used as the TOL. Where combinations of measures are used, for comparisons across studies the constituent parts of the score should also be provided (Berg & Byrd, 2002). In the next section we will consider the way in which tower-based problems may be solved.

1.2.2.4 Strategies: algorithms and heuristics

The information-processing approach provides several descriptions of the way in which people attempt to solve problems. This can be achieved using algorithms or heuristics (Unterrainer & Owen, 2006). An algorithm searches all possible solutions and will always find the goal, so it is used when there is enough capacity for doing so (i.e., by computers) but it is an inefficient method because it is an exhaustive search. Because humans do not have the capacity for using algorithms they perform selective searches, or heuristics, which are more efficient but not always successful in reaching the goal (Unterrainer & Owen, 2006). Some heuristics used by humans are termed: trial and error, hill climbing and means-end analysis (Dunbar, 1998). Trial and error approaches involve choosing the next move at random. Hill-climbing (sometimes called a perceptual strategy, or difference-reduction) is a more sophisticated method that involves a comparison of one's current state to one's goal state, and taking steps, one at a time, to reduce the difference between the two. However, this method can be misleading: Unterrainer & Owen (2006) call it short-sighted. This can occur when the optimal next move temporarily takes the solver *away from* the overall goal (sometimes called a counterintuitive move). Thus, someone using the hill-climbing method would run into difficulties in this type of problem (an illustration of this type of problem can be seen in Figure 1.3: the first move should be to move the red piece to the centre, even though it is already on the left hand peg – its 'goal peg'). A method that can be used in this type of situation is called means-end analysis, or subgoaling (confusingly, some authors also use the term means-end analysis when describing difference-reduction, e.g., Kahney, 1986b). It was a feature of the GPS model, proposed to characterise much of human problem solving by Newell and Simon (1972), and involves the setting of subgoals on the way to the goal. If a subgoal cannot be reached, a more intermediate one is set up, and flexibly solved, until the final goal is reached (Unterrainer & Owen, 2006). Because of working memory demands, Polson (cited in Kahney, 1986a) argues that it is not possible to plan the whole sequence of moves to solve a problem. Also, because people do not have a total understanding of the whole structure of a problem, they may be unable to set appropriate subgoals. He suggests that moves are generated as one goes along, using working memory and means-end analysis. The model consists of an adapted form of means-end analysis and takes into account limited memory capacity.

How does the ability to solve problems develop? Some experimental work has described children's problem solving on tower-based problem solving tasks, and is reviewed below.

1.2.2.5 Problem solving in typical development

Piaget (1976) described children's solving of the TOH. Several stages of ability were identified: children at stage 1 did not plan or organise their actions and instead engaged in trial and error, having difficulty with 2-disc problems; for children at stage 2, prediction and trying to move from the means to the ends were said to characterise this stage, while they solved the 2-disc problem without difficulty and the 3-disc tower with some errors. Behaviour was more directed towards the end goal than in stage 1. Stage 3 children (11-12 years) were quickly able to complete the 3-disc tower and could use their experience to transfer this to problems with more discs.

Klahr and Robinson (1981) adapted Simon 's Tower of Hanoi (TOH) task so that it was suitable for children, aged three and a half to six. Children were asked to verbally plan their moves to move a family of monkeys from one tree to another, without physically moving any pieces, on tasks varying from 1-7 moves (for the 1-move problems the experimenter carried out the child's instructions; thereafter, the 'pure planning' mode meant that no pieces were moved). Tasks varied by goal state: they were tower-ending (all monkeys sitting on top of one another) or flat-ending (all monkeys in their own tree). In the tower-ending tasks, two thirds of 5-year-olds and most 6-year-olds could plan four moves, and over half the 6-year-olds, six moves. Four-year-olds could manage 2-move problems but had difficulty with longer solutions. Flat-ending tasks were more difficult than tower-ending tasks.

Klahr and Robinson (1981) and Spitz, Webster and Borys (1982) found that participants displayed worse performance on tasks where the sequence of subgoals required was ambiguous: e.g., flat-ending problems. Klahr (1985) investigated young children's (45 to 70 months) performance on a problem solving task that, in the same way, did not give inherent clues about the sequence of subgoals required. The Dog-Cat-Mouse (DCM) puzzle involved a box with grooves along the top, the three animals and some cheese, a fish and a bone. The food was fixed but the animals could move along the grooves to a corner. The aim was to match the animals to their preferred food. Generally children could plan and carry out a perfect solution from 2-3 moves away. They were sensitive to 'partial evaluation', that is, if one of the animals was in the right spot they were reluctant to move it away again in the search for the solution. This sensitivity was associated with poorer performance on the tasks. The authors concluded that strategies used were not trial and error because double (backup) moves were generally avoided and optimal paths were followed once the goal had been identified, also because of the use of the aforementioned partial evaluation.

It has also been found that when 3-year-olds are approaching a problem they will make a direct movement to reach a goal, but will break the rules of the game if their route to a subgoal is blocked. Even by age 6, counterintuitive moves temporarily away from the goal are difficult (Siegler & Alibali, 2005) and this is the type of move that is required in the context of Klahr's sensitivity to partial evaluation. In other tasks, problem solving also becomes more efficient with age: for example, in route planning, there is more backtracking seen in 4-year-olds than in 5-year-olds (Fabricious, 1988, cited in Siegler & Alibali, 2005).

Thus, even early problem solving develops with age, and children find problems difficult when the order of the subgoals is ambiguous and when a temporary move away from the goal state must be made. This speaks to the strategies introduced in Section 1.2.2.4: difficulties with counterintuitive moves are likely to indicate a difference-reduction, (hill-climbing) strategy, as if one cannot look ahead, one will not be able to plan to remove and then return the piece to make the counterintuitive move. However, counterintuitive moves have also been found to affect adults' performance (Phillips, 1999), and indeed perceptual strategies have also been suggested to be used by adults (Goel & Grafman, 1995). Performance on tower-based problem-solving tasks improves over developmental time (Anderson et al., 1996; Huizinga et al., 2006; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Luciana & Nelson, 1998).

What drives the development of problem solving? It is recognised in the information-processing literature that an individual's skills do not suddenly appear in adulthood, and must somehow be developed over time (e.g., Anderson & Lebiere, 2003). Computational models have been proposed which address the development of children's abilities to solve problems on tasks other than the TOL:

for example, balance scale problems, in which different combinations of weights are added to a balance scale at different distances from the centre and children try to predict the outcome (see, e.g., Jansen & van der Maas, 2002; van Rijn, van Someren, & van der Maas, 2003). Such models attempt to identify rules which describe children's behaviour (see, e.g., Siegler, 1983 for a discussion of rule-based theories of cognitive development). 'Intelligent tutor' programs have also been developed to understand and respond to a student's problem solving for use in algebra lessons (Anderson, 2012). Information-processing accounts of developmental improvements on the TOL task are much rarer. However, Baughman and Cooper (2007) have attempted to account for such development using a Production Systems, rule-based approach, in terms of improvements in an aspect of executive functioning (EF) (see Section 1.2.3.4). EF is a somewhat separate body of research, although it has been considered in relation to problem-solving literature more recently. Goal-directed behaviour (i.e., problem solving) is thought to rely on executive functions (EFs) (e.g., Banich, 2009), and it is to these that we will now turn.

1.2.3 Executive functioning (EF)

EF is thought to include such abilities as planning, inhibition (the ability to withhold a prepotent response), shifting (the ability to switch between different mental sets), working memory (the ability to keep information in mind, and sometimes, to manipulate it as well), attention, fluency, monitoring, reasoning and goal formation (e.g., Banich, 2009; Best & Miller, 2010; Jurado & Rosselli, 2007). Today's EF research has its roots in both neuropsychology and cognitive science (cf. Baughman & Cooper, 2007; Shallice, 1982). Tasks that were found to be related to the functioning of the frontal lobes have been called 'executive', 'supervisory' or 'control' tasks, and have taken on their own discipline of behavioural EF research (Stuss & Alexander, 2000). EF tests have been found to activate several areas of the frontal lobe, including areas of the PFC: see Jurado and Rosselli (2007) for a list of studies.

1.2.3.1 Issues in EF research

In this section, pertinent issues in interpreting EF research are reviewed that are discussed by various authors (e.g., Banich, 2009; Jurado & Rosselli, 2007; Stuss & Alexander, 2000). The first concerns the way in which EF is conceptualised. There is a lack of agreement regarding whether EFs should be treated as a single entity or a collection of related components, known as the unity/diversity problem, while alternative approaches have also been taken, e.g., attempting to represent EF as separate phases of problem solving (representation, execution and so on) (Zelazo, Carter, Reznick, & Frye, 1997). Some accounts of cognitive control postulate the existence of one overall construct that governs behaviour (e.g., Duncan, Emslie, Williams, Johnson, & Freer, 1996; Shallice & Burgess, 1991b's SAS). The central executive in Baddeley's model of working memory (e.g., Baddeley, 1986) is another oft-cited candidate. Conversely, EF is often represented as a collection of processes that underlie goal-directed behaviour (Lehto et al., 2003; Miyake et al., 2000). This will be the approach adopted within this thesis. In a seminal paper, Miyake et al. (2000) measured EFs in typically developing adults, using confirmatory factor analysis (CFA) to identify shared constructs, and suggested that there are three related but distinct EF components: shifting, inhibition and monitoring/updating (manipulation and processing - rather than simple storage - in working memory). Miyake and colleagues suggested that the EF components could be related because of common demands across tasks: their potential candidates were 'controlled attention' (Engle et al., 1999a,1999b, cited in Miyake et al., 2000), such as maintaining a goal or resisting distractions, or simply inhibition as being necessary for all EFs to occur. Low correlations between EFs are also seen in typical adults, adolescents and individuals with brain damage, as well as children (Lehto et al., 2003), and Godefroy et al. (1999) provide evidence for multiple control processes in the brain that subserve EFs, which further supports the existence of a collection of moderately related functions (Jurado & Rosselli, 2007).

The second issue in EF research is that so-called 'frontal' and 'executive' functions are often (inappropriately) seen as equivalent. This is because of the ill-defined relation between frontal function in the brain and EFs measured using cognitive tasks (Stuss, 1992). With this in mind, EFs will be considered on a cognitive basis, but with the understanding that the frontal lobe, at least in typical adults, is the main brain area attributed to such functions.

The third consideration is the task impurity problem (Burgess, 1997; Phillips, 1997, cited in Miyake et al., 2000): by their nature, EF tests are complex, and thus depend upon (and therefore measure) many underlying skills. For example, inhibition and working memory underlie shifting (Davidson, Amso, Anderson, & Diamond, 2006). So, even if many factors are found to contribute to EF, or if EF task performance does not correlate with another measure, this could be due to differences in lower level skills (e.g., language) on which the task also depends, rather than differences in EF *per se*. Therefore, it is conceivable that different clinical groups might fail on the TOL for entirely different reasons: their performance might look the same, but the cognitive processes driving that performance might differ (Stuss & Alexander, 2000).

Fourth, EF tests have low test-retest reliability, which could be due to the effect of repeating a task, meaning that it becomes less novel, and perhaps less executive, at retest (Rabbitt, 1997).

Finally, the ecological validity of EF tests has been challenged: empirical tests have been argued to have poor representativeness and generalisability with respect to everyday functioning (Burgess et al., 2006). Because EF and the PFC is dependent on the outputs of several other systems, i.e., can be seen as 'central', (in line with the unity approach described above) Burgess and colleagues argue that measuring EF should be considered as different to measuring a function that is more basic and therefore more easily translatable to the real world, such as visual perception or auditory processing. They point out that while "models of the world" are usually used to bring the real-world requirements into an experimental situation (e.g., motion detection), EF tasks like the Wisconsin Card

Sorting Test (WCST; D. A. Grant & Berg, 1948) requiring flexible card sorting (see Burgess et al., 2006 for a discussion of its history) do not seem to represent the outside world in the same way. The WCST "is so unlike everyday situations, that knowledge of performance in it is of very little help for assessment since it is of uncertain predictive validity ("or generalizability")" (Burgess et al., 2006 p.199). They argue for real-world functioning to be taken into account when developing tests of EF. Various responses to the ecological validity problem have been developed (Spooner & Pachana, 2006) which are described in Chapters 4 and 6.

1.2.3.2 Issues in developmental EF research

Studying EFs developmentally gives the opportunity to assess time-based sequences and identify interactions, for example, two EF skills facilitating another skill, or a consolidation period in which skills are integrated (Garon et al., 2008, cited in Best & Miller, 2010). Some additional issues apply to developmental EF research. In a review article, Best and Miller (2010) highlight several relevant considerations and challenges which are discussed below.

There is evidence of the same related yet distinct system of EFs in children as in adults (Huizinga et al., 2006; Lehto et al., 2003), although the patterns seem to differ in childhood (Best & Miller, 2010). This illustrates the importance of considering children's abilities as they are developing, rather than assuming that the phenotypic endstate is constant throughout development (Karmiloff-Smith, 1998).

Another characteristic of developmental EF work, pointed out by Best and Miller (2010), is based on the truism that as children get older they become more able. Therefore, it seems that tasks must either become more complex, thus involving more EFs to provide a challenge (related to the task impurity problem, above), or that different tasks must be used at different ages, which makes comparisons difficult. Moreover, even when the same task is used, children of different ages might fail, or succeed, for different reasons.

To summarise, there are several issues that should be considered when engaging in EF research generally, and also developmentally. The way in which EFs are conceptualised is not clear cut, but a set of associated but separable constructs tends to be agreed upon, although this is by no means unanimous. The structure of EFs will not necessarily remain constant across developmental time. While EFs are thought to be located to the prefrontal cortex (PFC), the mapping is not equivocal. In measuring EFs developmentally, we must not only consider the task impurity problem but also the relative task demands and how they interact with children's age-related abilities. The ecological validity of EF tasks has not been reliably established.

1.2.3.3 The development of EF

Very early in development, success on detour reaching (around a barrier to grasp an object) and A-not-B tasks (being able to update the reached-for location when a previously reached-for object is visibly moved) demonstrate emerging goal-directed behaviour (Davidson et al., 2006). Rapid improvements in inhibition and cognitive flexibility are seen in 3-5 year olds (Davidson et al., 2006). Changes in EF have been linked to changes in the brain. For example, the localisation of inhibition circuits in the brain increases over developmental time, and the PFC undergoes protracted development, not reaching maturity until late in adolescence (O'Hare & Sowell, 2008, cited in Best & Miller, 2010). There are several types of inhibition task, including simple (stopping a response), complex (stopping a response and making a different response, also involving working memory) (Garon et al., 2008, cited in Best & Miller, 2010), and conflict tasks (Carlson & Moses, 2001), the latter of which demand an opposing response, such as saying 'day' to a picture of the moon, in Diamond's day-night task (Best & Miller, 2010). Children improve at complex inhibition tasks rapidly over the first few years of life (Best & Miller, 2010). There is some evidence of later improvements, generally on computerised tasks, such as a Go-NoGo task, in which participants respond to several types of stimuli, but must withhold their response to one type, e.g., a red one (Best & Miller, 2010). They note that inhibition improves most in the first few years of childhood, with shifting and working memory showing a more protracted course of development.

Working memory (cf. Baddeley's working memory model, Baddeley, 1983) comprises verbal and visuospatial subsystems: the phonological loop and the visuospatial sketchpad respectively. It can also be divided into simple (storage) and complex (storage plus processing) tasks, which also recruit the central executive control system (Gathercole et al., 2004, cited in Best & Miller, 2010). Working memory capabilities increase throughout childhood (e.g., Huizinga et al., 2006; see below).

Some simple shifting tasks can be completed by 3-4 year olds (Hughes, 1998, cited in Best & Miller, 2010) such as the Dimensional Change Card Sorting task (DCCS; Zelazo, 2006), a simplified version of the WCST. Based on DCCS performance, Zelazo et al. (2003) proposed a theory of children's early EF development in which children silently form plans in the shape of rules, which govern their internal and external behaviour on a task; rule systems can become more complex with age and are reflected on more, which account for increases in EF over developmental time. There is also evidence for increased shifting abilities with age (Huizinga et al., 2006; Luciana & Nelson, 1998).

Best and Miller (2010) discuss some proposed developmental sequences in the literature, as outlined below. Anderson et al. (2002) suggested that attentional control (inhibition), information processing (processing speed), cognitive flexibility (switching), and then goal setting develop in that order; Romine and Reynolds (2005; cited in Best & Miller, 2010) suggest a developmental sequence of inhibition of perseveration, followed by set maintenance, design fluency, planning, and then verbal fluency; while Carlson (2005; also cited in Best & Miller, 2010) in assessing the probability of passing EF tasks at different ages, suggested that inhibition matures at the earliest time point, then working memory, followed by those tasks that involve both of these skills. Stuss (1992) conceptualised three levels of operations: sensory-perceptual, executive functioning, and self-reflectiveness. He reviewed biological and psychological data about the development of these abilities and noted parallels between them and the model proposed: e.g., brain areas serving functions at the lowest level in the model developing before those in the EF levels.

In the next section, studies are reviewed in which EF skills have been measured alongside tower-based problem-solving performance, bringing us closer to understanding the abilities which might underlie problem solving.

1.2.3.4 Which EF abilities are important for TOL performance?

Numerous studies have examined contributions of EF tasks to performance on tower-based tasks, but as differences in the task contributions between the TOL and TOH have been found, both in children and adults (Bull et al., 2004; Welsh et al., 1999; Zook, Davalos, DeLosh, & Davis, 2004), research regarding the TOL will be the primary focus here. The TOL is a complex task involving the integration of different processes (e.g., Berg et al., 2010; Koppenol-Gonzalez, Bouwmeester, & Boonstra, 2010). In typical adults, evidence regarding the reliance of TOL performance on working memory is somewhat mixed, with visuospatial memory emphasised by some authors, verbal rehearsal by others, and one study finding no evidence (see Unterrainer & Owen, 2006 for a review). Welsh et al. (1999) suggested that working memory and inhibition were important for TOL performance, while Unterrainer et al. (2004) found that nonverbal fluid intelligence predicts TOL performance in adults, and that a 'planning' factor was the only construct explaining performance. The role of planning was challenged by Phillips (1999); Phillips, Wynn, McPherson, and Gilhooly (2001).

The ways in which EFs relate to TOL performance in typical children will now be addressed. Some studies investigate this in one age group (e.g., Bull et al., 2004) while others take a developmental approach, administering several tasks to children over a range of ages (e.g., Huizinga et al., 2006). Across studies, results have been used to identify the factor structure of EF and to examine patterns of age-related EF development. The studies reviewed here have included tower-based problem-solving tasks in their task battery, often as a 'complex' EF task, with attempts made to relate lower-level EF tasks (e.g. inhibition) to more complex tasks. Thus, from a problem-solving point of view, they provide insights into performance with increasing age, and into the relationship with EF skills.

Anderson et al. (1996) found significant relationships between the TOL and a range of tests of EF measuring mental flexibility and planning, abstract thought and organisational ability, in a sample of children aged 7-13 years. There were notable increases in performance on the TOL between 7- and 9-year-olds, and at around 11 years. Lehto, Juujärvi, Kooistra, and Pulkkinen (2003) administered a battery of EF tasks to 108 children aged 8-13 years, split into five age groups. Age improvements were found on 9 of the 14 tasks, including the SOC, and the Working Memory and Shifting factors were related to CA. Shifting, inhibition, auditory attention (which also involved shifting), word fluency, a maze task and spatial working memory were correlated with TOL performance measures. They identified three EF factors (Working Memory, Inhibition and Shifting), and moderate correlations between EF tasks, as well as significant correlations between factors after controlling for chronological age (CA), supporting the conceptualisation of related but distinct EFs seen earlier (e.g., Miyake et al., 2000).

Several authors have proposed that inhibition has a role to play in TOL performance, with some suggesting that it is important for age-related improvements. Asato, Sweeney and Luna (2006) administered SOC problems with 2, 3, 4 and 5 moves, to TD individuals aged between 8 and 30 years, split into groups of 8-13, 14-17 and 18-30 (children, adolescents and adults). They measured inhibition, spatial working memory and information-processing speed using occulomotor (eye-movement) tasks. Performance on each task improved with age. There was a significant positive correlation between better inhibition scores and better SOC performance (making fewer moves, as well as timing-related measures) on the more difficult problems, while only modest correlations were observed for the working memory measure. The authors emphasise the role of inhibition in SOC performance, and suggest that the ability to inhibit responses is relevant to its development. Huizinga, Dolan and van der Molen (2006) gave groups of TD participants aged 7, 11, 15 and 21 years various tests to measure working memory, inhibition and shifting abilities as well as the WCST and TOL. Performance increased with age on all tasks. Of their TOL measures, the number of additional moves and planning time reached adult levels at 15 years of age. The proportion of perfect solutions continued to increase between each age group up to the young adults. Stroop performance, measuring inhibition, predicted the percentage of perfect solutions in adults, while this analysis was inconclusive for the children, but the use of different strategies in children and adults was suggested. Baughman and Cooper (2007) modelled young children's TOL performance (including rule breaks) for 3-4 and 5-6 year old children, and suggested that the emerging ability to inhibit a perceptual strategy is an important factor in accounting for development. Luciana and Nelson (1998) presented the CANTAB test battery to TD children aged 4, 5, 6, 7, 8 years as well as to teenagers and adults. They varied the number of moves required to complete a problem (2, 3, 4 or 5). While all groups made some excess moves, 4-year-olds made more in 3-move problems than the other groups, while the remaining groups of children made more in 4- and 5-move problems than the adults did. The proportion of participants who solved problems in their minimum number of moves appeared to generally decrease with the number of moves for all the younger age groups, while the 4-year-olds appeared to have more difficulty earlier (at 3 moves) than the older children. Luciana and Nelson found age-related improvements on executive tasks, but did not correlate performance between tasks. However, they noted that working memory performance paralleled SOC performance for 4-year-olds, suggesting that working memory abilities influenced children's problem-solving strategies on this task.

For some studies, interpretation of which EFs are related to TOL performance is bound up with the properties of the task (described in Section 1.2.2.3): that is, different types of TOL problems are thought to rely on different EF skills. For example, Kaller, Rahm, Spreer, Mader, and Unterrainer (2008) measured TOL performance (on a variant of the TOL in which all pegs were the same height) in 4-5 year-old children. All of their TOL problems were 3-move problems, but they varied in their other parameters: that is, whether or not they contained the requirement to make an intermediate move (i.e., search depth), as well as the goal hierarchy. They also administered tasks measuring inhibition (Go-NoGo) and working memory (forwards verbal and visuospatial span tasks). While inhibition response time to 'Go' trials and verbal (not visuospatial) STM predicted overall TOL accuracy, only CA accounted for the difference in accuracy on problems with different search depths. Kaller and colleagues interpreted this as an age-related increase in the ability to search ahead when solving these problems. They refined Luciana and Nelson (1998)'s suggestion that their 4-year-olds found 3-move problems challenging in terms of their limitations with using working memory, as difficulties with planning ahead for that intermediate move, as Luciana and Nelson's 3-move problems all required an intermediate move. Bull, Epsy and Senn (2004) gave groups of 4-year-old children either a TOH or a TOL task in which problems varied from 1 to 7 moves, and required 0, 1, 2, 3 or 4 moves involving conflict between goals and subgoals, which they call counterintuitive (it should be noted that some of these counterintuitive moves temporarily blocked another ball's move to the goal rather than moving in the direction away from the goal peg; nevertheless they all involve a conflict between the goal and the subgoal). Shifting and Inhibition were also measured (from 'The Shape School' executive functioning task; Espy, 1997) involving naming shapes, inhibiting naming shapes with sad faces, then shifting between naming shapes by their colour or shape, depending on whether or not they wore a hat. Short term memory (STM) was measured by children's maximum forwards digit span. TOL performance was correlated with CA, shifting accuracy, and inhibition time, although when the baseline speed of shape naming was controlled for, the correlation between inhibition time and TOL performance ceased to be significant. As the number of counterintuitive moves in a problem increased, 4-year-old children's ability to shift between tasks and inhibition responses increasingly contributed to the variance in performance in the TOL. This increased contribution became significant at two counterintuitive moves. This study suggests that for 4-year-olds, shifting and inhibition (but not working memory) become more important for performance as the number of counterintuitive moves in a problem increases.

Some characteristics of the studies should be taken into consideration when examining contributions of EF skills to problem solving. The particular tasks which are administered alongside the TOL are important for interpretation. For example, both Bull et al. (2004) and Kaller et al. (2008), while not finding a contribution of working memory performance to the TOL task, noted that they have only included measures of STM, as opposed to executive working memory, which may be more relevant in a task like the TOL. Also, variations in scoring and administration procedures in the TOL should prompt caution when interpreting outcomes or comparing across studies (Berg & Byrd, 2002). For example, Kaller and colleagues' (2008) study uses the Ward and Allport (1997) variant of the TOL. This version is substantially different from the 'Shallice TOL' because being able to place any piece on any of the pegs (unrestricted by height) means that the problem space is altered, as well as the cognitive demands on the individual, as they do not need to remember the rules (Berg & Byrd, 2002). In addition, some studies have required participants to plan their moves before they start, and/or only allowed problem success when the minimum number of moves is produced perfectly (e.g., Klahr & Robinson, 1981; Luciana & Nelson, 1998). Knowing how many moves are required beforehand is likely to increase the amount of planning that is engaged in (Berg & Byrd, 2002) so being restricted to that number would presumably have an even stronger effect.

From these studies, we can see that a) EF can be divided into several related factors in typical children (as well as in adults: see Section 1.2.3.5) and b) that several EF skills, and perhaps inhibition in particular, are related to TOL performance in children. EF abilities are impaired in ADHD and autism (Pennington & Ozonoff, 1996) as well as other clinical groups, such as Parkinson's disease and schizophrenia (Unterrainer & Owen, 2006) and including WS and DS (see below), and executive (but not TOH) impairments were also seen in adults with general intellectual disabilities (Danielsson, Henry, Rönnberg, & Nilsson, 2010). While the contribution of EFs to TOL performance has been explored in TD, we do not yet know how those EF skills might be used to solve problems in neurodevelopmental disorders. In the next section, what is known about EF and problem solving in WS and DS is reviewed. Children with learning difficulties have been found to be impaired on a TOH task, performing at the level of younger children, and also using different strategies (Spitz et al., 1982). Therefore, it is likely that individuals with WS and DS will also show impairments in problem solving. It should be noted that although relative strengths and weaknesses in EF profiles are investigated and discussed, this is always within the framework of overall diminished EF abilities for their age, demonstrated, whether explicitly acknowledged or not, through any study in which performance of an atypical group is poorer than or equivalent to a matched TD group with a lower mean CA, e.g., a mental-age (MA) matched group.

1.2.3.5 Executive functioning and TOL performance in WS and DS

1.2.3.5.1 Williams syndrome

In WS, working memory abilities, and EF abilities more generally, are significantly below the level that would be expected for chronological age (CA) (Osório et al., 2012; Sampaio, Sousa, Férnandez, Henriques, & Gonçalves, 2008). The uneven cognitive profile in WS (Mervis et al., 2000) is generally reflected in the group's relatively better performance on verbal short-term memory (STM) tasks compared to spatial STM tasks (e.g., Jarrold, Baddeley, & Hewes, 1998; Wang & Bellugi, 1994), while verbal working memory has been found to be below the level expected for receptive vocabulary, and spatial working memory at or below the level expected for nonverbal reasoning ability, dependent on the type of memory measured (Rhodes, Riby, Fraser, & Campbell, 2011). Visual-spatial (but not visual-object) STM and long-term memory (LTM) is also impaired in WS compared to MA (Vicari, Bellucci, & Carlesimo, 2005; Vicari & Carlesimo, 2006).

There is evidence that the better verbal than spatial abilities seen in WS also extend to inhibition (while acknowledging that the verbal/spatial distinction is a broad one; see Section 1.2.1.2 and Brock (2007) for an outline of language abilities in WS; also, e.g., Pezzini, Vicari, Volterra, Milani, and Ossella (1999) for evidence of strengths and weaknesses within both cognitive and linguistic domains). Atkinson et al. (2003) compared the inhibition abilities of children with WS aged 4:7 to 14:7 years to a control group aged 4:1 to 13:5 years on a task in which they were asked first to point to a target when it appeared on the screen (pointing task) and subsequently to point to the other side of the screen as the target when it appeared (counterpointing task). The children with WS performed more poorly on the counterpointing task relative to the pointing task than did the controls, indicating difficulties with spatial inhibition. On a detour box task requiring inhibition of a previously learnt strategy to retrieve a toy, and a day-night task requiring saying 'day' to a card depicting the night sky and 'night' to a picture of the sun during the day, larger groups of participants showed delayed performance on the detour box task: while an improvement in performance is typically seen at approximately 3 years of age, for the WS group it improved at a verbal (BPVS) mental age (MA) of about 7 years. Verbal inhibition, as measured on the day-night task, was similar to or above the level expected for their BPVS scores. In contrast, Tager-Flusberg, Sullivan, and Boshart (1997), in a study investigating the relationship between EF and theory of mind, found that a group of children with WS performed at an equivalent level on a motor and verbal inhibition task. However, the sensitivity of the measures was poor, with participants awarded either a pass or fail for each task, based on a cut-off score. Of those who could complete both tasks, 10 out of 12 children with WS either passed both or failed both inhibition tests. Unfortunately a typical control group was not included in the study, although the WS group did not score differently from a group of children with Prader-Willi syndrome.

In some studies, individuals with WS completed a battery of tests which included a TOL task as a measure of planning. Menghini et al. (2010) compared a WS group ranging from 10 to 34 years of age to a TD group matched on mental age measured by the Leiter-R (Roid & Miller, 2002). This is a nonverbal test, thought to assess fluid intelligence. They found that auditory sustained attention, but not selective attention, was poorer in the WS group, while this pattern applied to selective but not to sustained attention, in the visual-spatial domain. In memory tasks, while non-word repetition was as good in the WS group as in the control group, (and was also found by Grant et al., 1997 to be comparable to nonverbal MA) digit and block memory tests (in which both forwards and backwards conditions were measured) were poorer in the WS group than for the controls, with the same pattern of poorer backward than forwards memory in both groups.

Categorising objects was poorer in the WS than the TD group, while a test of verbal fluency, included as the verbal counterpart to the visual-spatial categorisation test, did not reveal group differences. Impairments were seen on visual-spatial, but not verbal, shifting tasks. For measures of inhibition, verbal inhibition on the Opposite World test from the Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999) took longer in the WS group than for the controls, while the Stroop test (Stroop, 1935) did not show group differences in interference scores based on RT (although there was a trend towards more errors in the WS group for the incongruent condition). In a Go-NoGo test assessing visuospatial inhibition, the participants with WS made more errors as a group, but did not take longer than the TD group. In a TOL test, the WS group took more repeated attempts to solve problems and solved fewer items correctly, but did not take longer than the TD group on correctly solved problems. The authors draw on the parallels between inhibition and TOL tasks in that similar speeds between groups but more errors in the WS group were seen in both areas (with the exception of the Opposite World test), emphasising the role that impulsivity might play in the way in which participants with WS solve problems on the TOL. They also point out that these two abilities (planning and inhibition), along with working memory, show impairments across modality, while the shifting, categorisation and attention tasks were more modality-dependent.

Rhodes, Riby, Park, Fraser, and Campbell (2010) administered the computerised CANTAB EF test battery (Sahakian & Owen, 1992) to a group of individuals with WS (mean CA: 18 years 1 month; mean BPVS raw score: 92.95), chronological age matched controls (CM group; mean CA: 17 years 5 months; mean BPVS raw score: 126.79) and verbal age matched controls (VM group; mean CA: 9 years 3 months; mean BPVS raw score: 92.26). Compared to VM and CM groups, WS performance was impaired on a set-shifting task and on several working memory tasks: one designed to tap executive aspects of working memory, involving 'searching' an array of boxes for blue tokens and avoiding returning to the same box; one measuring a memory span; and another requiring the matching of a visual stimulus with and without a variable delay. The Stockings of Cambridge (SOC) task was also administered, which is both similar to the TOL task in some ways and different in others (Berg & Byrd, 2002; Berg et al., 2010). Problems varied by the minimum number of moves required to reach the solution (2, 3, 4 or 5). The WS group solved significantly fewer problems in the minimum number of moves than CM and VM groups. Also, while the groups made comparable numbers of moves for problems with two moves, the WS group made more moves than both the other groups on problems with three moves, and more than the CM group for 4- and 5-move problems. Performance levels on different tasks were associated in the TD group, but not for the WS group.

In summary, there is evidence that individuals with WS experience difficulties with EFs relative to what would be expected for their MA, as well as some evidence for greater impairments in visuospatial than verbal modalities.

1.2.3.5.2 Down syndrome

Working memory abilities have been widely studied in DS. Broadly, the opposite pattern to that seen in WS has been observed, with better visuospatial than verbal STM (e.g., Jarrold & Baddeley, 1997). Several studies have administered a range of EF tasks to individuals with DS, matching groups on various measures of overall MA, verbal ability and logical thinking. Pennington et al. (2003) compared prefrontal skills, as well as general functioning measures and hippocampus measures, between adolescents with DS and a control group of TD children matched on scores on the Differential Ability Scale (DAS; Elliott, 1990). No significant differences between groups were found on the SOC task from the CANTAB battery, or indeed for any of the other prefrontal tasks: the spatial working memory test from the CANTAB, the NEPSY verbal and design fluency tasks, Stopping (inhibition) task and Counting Span (verbal working memory) task. In contrast, the DS group performed more poorly than MA controls on measures of hippocampal function. Although they acknowledge complexities in the design and interpretation, they suggest hippocampal, rather than prefrontal (pertaining to EF tasks) dysfunction in DS. Rowe, Lavender, and Turk (2006) administered several EF tests to a DS group (aged 23-40 years) and a BPVS-matched group with other learning disabilities (aged 19-55 years). With the exception of measures of verbal fluency and verbal STM (digit span), the tasks were nonverbal: measuring sustained attention, motor perseveration/inhibition, Raven's Coloured Progressive Matrices (RCPM; Raven, Court, & Raven, 1990), spatial STM, TOL and set-shifting, as well as a measure of motor speed. The two groups did not differ reliably on the spatial span task performance while for the other tasks the DS group scored more poorly than the learning disabled group, with shifting, sustained attention and RCPM differences showing the strongest discrepancies. The motor speed test also revealed slower responding in the DS group.

Lanfranchi, Jerman, Dal Pont, Alberti, and Vianello (2010) compared EF in 15 adolescents with DS between the ages of 11:0 and 18:5 (mean age = 15:2), with that of a TD control group with a mean CA of 5:9 (4:6 to 6:10). Participants were individually matched with TD controls on MA, the mean of which was also 5:9 (4:6 to 6:10) for the DS group. MA was measured using the Logical Operations test (Vianello & Marin, 1997), assessing seriation, numeration and classification abilities, which was chosen for its relatively reduced susceptibility to extraneous influences (i.e., verbal, cultural or visuospatial). Adolescents with DS performed significantly more poorly than the MA-matched TD group on the majority of the EF tasks. In a verbal and a visuospatial working memory task demanding dual-task processing, the DS group was less able to remember the first item in a list or starting position of a character as well as to tap the table when a particular event occurred (hearing the word 'ball' in the verbal task and a frog jumping onto a red square in the visuospatial task) than the TD group. They were poorer at a day-night task than controls although they were just as able to correctly say 'day' or 'night' when the words were arbitrarily associated with an abstract image, indicating difficulties with inhibition of prepotent responses relative to MA. Set shifting was poorer than that of controls, in a task requiring participants to say 'Yes' and 'No' to different coloured cards: firstly Yes to red cards and No to black and subsequently Yes when a card was the same colour as the previous card and No when the colour had changed. While both groups were equally able to correctly respond to the first part of the task, the number of correct responses declined in the part of the task that followed the rule change for the DS group such that the number of correct responses was lower for DS than TD participants. On the Modified Card Sorting test (MCST; Nelson, 1976) measuring the ability to shift between concepts, the number of perseverative errors in sorting did not differ between the groups while the DS group was able to identify and follow fewer rules than the TD group. In a TOL task with a standard start state and 12 problems of increasing difficulty, the number of correctly solved problems was significantly lower for the DS group (M = 3.8) than for the TD group (M = 9.67). Participants were given three attempts to solve each problem. The number of problems solved in the first attempt, perhaps unsurprisingly, was also lower in the DS group (M = 2.67) than the TD group (M = 5.67). Group differences were not apparent on tests of verbal fluency while in a test of sustained attention, the Self-ordered pointing test (Temple, Carney, & Mullarkey, 1996), the DS group made more errors than the TD group. The authors noted that the most severe impairments in DS relative to MA were in verbal working memory, in shifting on the MCST and in the TOL. Verbal fluency was the only task that did not reveal DS impairments for their overall cognitive level. They also correlated CA with EF task performance, and found only two significant relationships: between CA and verbal working memory and between CA and TOL performance solely for the TD group.

Thus, the outcome of EF profile comparisons appears to partly depend on the group matching measure that is selected: in Lanfranchi and colleagues' study all tasks elicited impairments in the performance of the DS group with the exception of verbal fluency compared to a logical thinking matched group, while spatial STM ability was comparable to that of a verbal-matched group in Rowe and colleagues' study. Furthermore, in Pennington et al.'s study, no EF task deficits were found in the DS group in relation to overall MA. The modality of the task is also potentially important for interpreting outcomes, given the uneven EF profile in DS; however this rather mixed picture of EF abilities in DS does provide some evidence of EF impairments in relation to overall cognitive ability.

1.2.3.5.3 Williams syndrome and Down syndrome

Some authors have investigated EF in both WS and DS (more specific details regarding TOL performance, e.g., timing and rule violations, are included in Chapter 3).

Jarrold et al. (1999) administered forwards versions of block and digit span tasks to individuals with each of the disorders. The WS group were better at digit than Corsi (block) span performance, while the DS group did not show this difference. Digit span was better in WS than DS, while Corsi performance was comparable between the two groups. The WS and DS groups did not differ in nonverbal MA, measured by a block design task of visuospatial construction, which is known to be a marked weakness in WS. When controlling for MA, WS Corsi performance was worse than DS performance, while the DS group showed a trend for scoring more poorly than the WS group on a digit span task. Other studies have also suggested differential patterns of STM performance in WS and DS (Jarrold et al., 1999; Vicari & Carlesimo, 2006; Wang & Bellugi, 1994).

Vicari et al. (2000) included the TOL as an implicit memory test in an investigation into implicit and explicit memory. No significant effect of group was found between participants with DS (mean CA: 21 years; mean MA: 6.5 years) and TD children (mean CA; 5.09 years; mean MA: 6.3 years) matched for overall MA, on a TOL score based on the number of attempts required in order to reach a solution in the minimum number of moves (scores consisted of giving 3 points for problems solved in the first attempt, 2 for the second attempt and 1 for the third attempt). Both groups' performance improved to a similar extent after a one hour break, suggesting implicit memory abilities in DS that are comparable to those found in typical development. In a similar experiment with children with WS (Vicari, Bellucci, & Carlesimo, 2001), only the typical group improved, leading to a conclusion that procedural learning is impaired in WS, but not in DS. Although these two studies did not directly compare WS and DS groups, a main effect of Group, indicating lower TOL scores than typical controls, was found in the WS study but not in the DS study.

While most studies administering an EF task battery either consider WS or DS, two studies to date have included both a WS and a DS group; one which compares each group to controls (Carney, Brown, & Henry, 2013) and another which makes direct comparisons between atypical groups (Costanzo et al., 2013). Carney et al. (2013) investigated executive functioning in both WS and DS. They administered tasks designed to assess executive-loaded working memory, inhibition, fluency and set-shifting in both the verbal and visuospatial domains to individuals with WS and DS aged 8:1-18:11 and 10:4-18:11 years respectively, as well as a group of TD controls aged 5:0-8:0 years. There were 24 WS, 25 DS and 26 TD participants. Rather than comparing performance to a control group matched on a particular measure, they controlled for CA and MA using dummy-coded regression. Each atypical group was compared separately to the TD group, so direct comparisons between atypical groups were not made. IQ was measured using the Stanford-Binet Abbreviated Battery (Roid, 2003). Although this IQ test gives separate verbal and nonverbal IQs, it was the combined MA score which was used in the comparison to the TD group. The executive-loaded working memory tasks involved both processing and storage. In the verbal task participants were presented orally with a sentence and asked to decide whether it was true, and also recall the last word of one syllable in that sentence. In the visuospatial task participants judged which image out of three was the odd one out and then recalled its location on a blank board. Both tasks increased in difficulty by varying the number of sentences or arrays of three images that were shown, before the participant was asked to recall the relevant information. Both the WS and DS groups scored more poorly than the TD group on the visuospatial version of this task. The DS group also showed poorer performance than the TD group on the verbal version of the task, while the WS group did not. In the verbal inhibition task, in an initial 'copy' block, the experimenter said either 'doll' or 'car' and the participant repeated the word for 20 trials. This was followed by an 'inhibit' block whereby the participant was asked to produce the alternative word in response to the experimenter (i.e. 'car' if the experimenter said 'doll'). The task returned to another 'copy' and then another 'inhibit' block. Performance was measured in terms of time taken and errors made across the four blocks. The visuospatial version involved reproducing actions (pointed finger or clenched fist) instead of saying words. The WS group took longer to respond in the visuospatial inhibition task, and made more errors in the verbal task, than the TD group. There were no significant differences between DS and TD groups on either of the inhibition tasks. In the verbal fluency task, participants were asked to generate as many items belonging to a category (food/drink and animals) as they could in one minute, and the number of items produced as well as number of repetitions was scored. In the visuospatial fluency task participants produced as many unique designs as possible in one minute by joining dots (in a black dot condition, and another condition in which they had to connect only white dots). Again, the score was in terms of the number of designs and the number of repetitions of designs. The WS group was able to generate fewer unique designs in the visuospatial task than the TD group, while verbal fluency scores did not differ significantly. The DS group did not show significant differences compared to the TD group in either of the fluency tasks. In the set-shifting tasks, participants were required to switch between the categories as in the verbal fluency task, or between connecting empty or filled dots as in the visuospatial fluency task. Performance was measured by the switching cost (correct response cost and repetition cost) in relation to the fluency task score. Both the WS and DS groups had a greater verbal switching cost for repetitions than the TD group, than expected by verbal fluency score. Neither group exhibited more of a cost than the TD group in the visuospatial condition. The authors concluded that modality (the verbal/visuospatial nature of the task) has different effects in different EF domains, and that the verbal or visuospatial strengths that purportedly mark out WS and DS profiles affect the way in which people with WS and DS respond differently, depending on which EF ability is being measured. However, it should be noted that given the overall MA measure that was used to control for MA in the analyses, the results are difficult to interpret because of the very different cognitive profiles of these two distinct developmental disorders. Carney and colleagues asserted that because there are WS impairments in both modalities of the inhibition task, relative verbal strengths are not evident. The authors acknowledged that there may be different underlying reasons for poorer performance in the different modalities, as one relies upon error rate and the other on longer response time. Might the longer response time for the visuospatial task reflect the very uneven profile that the results overall seem to go against? Perhaps the relative verbal strengths in WS point to a more typical WS verbal system, such that copying a spoken word is relatively automatic and prepotent. If, in contrast, the motor/spatial system in WS is more atypical, perhaps the prepotent instinct to copy is less marked for this group. The longer response time might then reflect a deliberate thinking-through of each action. Moreover, if someone with WS is aware that they are likely to have difficulty with a task involving actions, this might affect the way they respond, eliciting more deliberate control than perhaps a verbal task to which they might feel more confident (and more impulsive) about responding to. Alternatively it may be that, as these results suggest, inhibition is a particularly pervasive deficit in WS which thus extends across modality, and the difficulties are reflected in different ways for another reason.

Costanzo et al. (2013) directly compared WS and DS groups (aged 10.7 to 34.9 and 8.6 to 21.2 years) and typical controls (aged 6.1 to 8.4 years) on several measures of EF. Controls were matched on MA as measured by the Leiter International Performance Scale-Revised, brief version (Roid & Miller, 2002), which measures nonverbal abilities (Glenn & Cunningham, 2005) although they had higher IQ levels than the WS or DS groups. Measures taken were auditory and visual sustained and selective attention, verbal STM and working memory (forward and backward digit span tests), visuospatial STM and working memory (forward and backward Corsi block span tests), planning (on the TOL), verbal and visual categorization, verbal and visuospatial shifting, and verbal and visuospatial inhibition. On the TOL, better scores were awarded for completing problems in fewer numbers of attempts. The score was lower for the WS than DS or TD groups, while the time taken was longer in the DS group than either of the other groups. Results for the other tasks indicated, relative to controls, impairments in the WS group on the visual selective attention, verbal sustained attention, backward digit, forwards and backwards block tasks, visual-spatial categorisation, and equivalent performance on visual-spatial attention, forward digit test, non-word repetition, verbal fluency/categorisation, verbal shifting, visual-spatial shifting, and Stroop (inhibition) tasks. For the DS group, impairments relative to controls were seen on visual selective attention, verbal sustained attention, forwards and backwards digit span, non-word repetition, forwards and backwards block span, visuospatial categorisation, verbal and visuospatial shifting, and Stroop (inhibition) performance, with equivalent performance to controls on visual-spatial attention and verbal fluency/categorisation. The DS group performed more poorly than the WS group on both verbal memory tasks and nonword repetition, on the Stroop task, and on one verbal and one visuospatial shifting measure (with the remaining visuospatial shifting measure showing equivalent WS/DS performance). No group differences were seen for visuospatial inhibition.

In summary, there is evidence of a range of EF impairments in WS and DS groups relative to MA-matched controls. Modality-specific profiles between the disorders are reflected more clearly in some domains (e.g., working memory), than in others, and mixed patterns of results seem to be partly elicited by variations in choice of matching measures and EF tasks.

Thus, while EF has been closely studied in WS and DS, investigations into the relative ways in which the two groups use EFs to solve problems have not been previously undertaken, and is thus one focus of this thesis. Indeed, investigations of this type are limited to Rhodes et al. (2010)'s assessment of associations between EF tasks in a WS group, from which no significant relationships emerged. Given the different WS and DS profiles of EF skills, we can expect them to take different approaches to problem solving (more detailed predictions are included in Chapter 2).

1.2.4 Problem solving outside the laboratory

Up until this point in the chapter, we have considered problem solving, and EF tasks, for which data are collected in an experimental situation. Of course, the vast majority of problem solving occurs in everyday life rather than the laboratory.

Making the metaphorical leap to the outside world invokes several overlapping issues simultaneously: that of well-defined versus ill-defined problems; routine versus novel tasks; and symbolic versus situated action (SA) approaches to cognition. These will be addressed in turn, but should not be considered to be independent of one another.

1.2.4.1 Novel problems and routine tasks

Some tasks are so well-practised that we seem to be able to complete them without paying attention to them, while arguably a task which is novel to the solver will require more effort to solve. Newell (1980) saw routine problems as those which are so well rehearsed that we no longer need to apply effortful control to solving them (Anderson, 1993). This distinction is also addressed by Norman and Shallice (1986; 1980); Shallice (1982, 1988) (cited in Shallice & Burgess, 1991b) in a model of cognition that builds on the Production Systems approach. Within this model, schemas (also called productions or routine programs) control a highly specialised action such as drinking from a cup, and are triggered by particular stimuli. As many different schemas can be triggered simultaneously, the selection of which schema to run is an important consideration. 'Contention scheduling' routinely selects from potential schemas based on whichever has the strongest trigger. In some situations, i.e., non-routine problems, however, more deliberate control is needed, and this is where the 'Supervisory Attention System' (SAS) is utilised, described as dealing with planning and biasing contention scheduling (for more detail, see Shallice & Burgess, 1991b). While behaviour under contention scheduling is fast, rigid and routine, behaviour under the SAS is flexible and slow. The TOL was developed in order to measure the functioning of the SAS: i.e., when tasks were novel. There are reports in the neuropsychological literature of patients with damage to the frontal lobes who can pass IQ tests in the laboratory, but are nevertheless unable to manage their everyday lives (Eslinger and Damasio, 1985; Milner, 1964; Stuss and Benson, 1986, cited in Shallice & Burgess, 1991b). Shallice and Burgess (1991b) suggest that their model's two levels of control can account for this (although by their own admission, not uniquely so). In their model, if the SAS is not functioning correctly, difficulties with planning would occur on novel tasks, while routine tasks would not be affected (Shallice, 1982). Shallice and Burgess (1991) describe evidence that Tower of London performance, but not other types of performance, was impaired for patients with damage to the frontal lobe, notably in a study by Owen and colleagues in which short-term memory performance did not significantly differ from that of the control group. Patients' slower responses during the task also led to the conclusion that the TOL was indexing (poor) planning ability for this group.

However, Shallice and Burgess (1991b) also described three patients for whom performance on cognitive tasks in a laboratory setting, including the TOL, were considered normal, who nonetheless experienced difficulties in everyday life as well as on scheduling tasks. This leads us to the subject of problem solving outside, as well as inside, the laboratory.

1.2.4.2 Well-defined and ill-defined problems

The TOH and TOL are both examples of well-defined problems, in which the solver is provided with a starting point (initial state), a goal to work towards (goal state) and sometimes, a means of reaching the goal (legal operators and operator restrictions; that is, actions that can be taken to solve the problem and constraints on how those operators are used) (Kahney, 1986a). The solver need only work within the clearly defined boundaries of the situation in order to solve the problem: all of the information needed to solve the problem is provided within the problem, such as what the solution will look like and what should be done to reach it (Pretz, Naples, & Sternberg, 2003). Well-defined problems have the obvious advantage of being suitable for study in experimental settings, and in allowing observation of behaviour under controlled circumstances.

In contrast, many problems are ill defined, in that at the beginning of the problem we may not know exactly what the goal looks like or what needs to be done before the problem will be solved (Pretz et al., 2003); for example, when buying a birthday present for a friend, several factors such as what will be suitable and how much it will cost need to be taken into account. The value of investigating problem solving in an ill-defined environmental arena is clear, as this is where individuals with neurodevelopmental disorders will require support to function effectively in their everyday lives: this approach has great ecological validity.

Although there are clear differences between well-defined and ill-defined problems, they could be argued to share some requirements, such as the need to keep focused on the goal or perhaps generate subgoals on the way to completion. Some researchers have downplayed the difference between well- and ill-defined problems, suggesting that people will go about solving them in similar ways (Dunbar, 1998). The distinction between well- and ill-defined problems seems to some extent to map onto the distinction between problem-solving performance in experimental and in everyday settings. While this is arguably not absolute (some experimental tasks could be ill defined; some everyday problems could be well defined) it seems that in the main, the problems that we solve in our everyday lives are the least well defined. Thus, studying problem solving in everyday settings requires the study of ill-defined problems.

1.2.4.3 Symbolic and situated action (SA) approaches

The information-processing approach (leading to tower-based task development) is part of a symbolic, cognitive science approach to human cognition, concerned with the mind and brain processes that occur within the individual (Norman, 1993). An example of a set of if/then rules used in a model for solving the TOL can be seen in Baughman & Cooper (2007). One difficulty with symbolic approaches is their perceived inadequacy for accounting for problem solving in the real world: the sheer amount of information available in all its complexity and chaos (Norman, 1993). The opposing theoretical standpoint known as situated action (SA) is advocated from the view that real-life cognition can only appropriately be studied in its real-life context, taking into account the interaction between the environment, culture, cognition and action (Norman, 1993). SA approaches advocate that there is no merit in separating different aspects, such as the individual from the context. Social interaction and culture play an important role for researchers with this viewpoint (Norman, 1993). However it should be noted that some authors have presented models of how symbolic systems, for example contention scheduling and the supervisory system, could operate (and fail) in everyday life (Cooper, 2002; Cooper & Shallice, 2000). Vera and Simon (1993) argue that SA approaches can be subsumed by symbolic approaches; that is, that symbolic approaches can explain and encompass what SA proponents claim that the approach adds.

Thus, in any consideration of problem solving that extends outside the laboratory, we must take into account the relative novelty of the task and how well defined (or otherwise) the problem in question is likely to be. This issue also prompts us to revisit the ecological validity problem of EF measurement, outlined in section 1.2.3.5. This subject is explored in Chapter 5, in which the relationships between experimental performance and parent-reported measures of everyday functioning are assessed. In short, while some studies have assessed the potential links between experimental and everyday measures of EF in the typical population and in clinical populations such as ADHD (e.g., Toplak, Bucciarelli, Jain, & Tannock, 2008) this type of investigation has not been undertaken in WS or DS groups. It is also unknown whether problem solving on the TOL is associated with real-life problem solving for these populations.

The aims of this thesis are thus to conduct a cross-syndrome comparison of the relationship between EF and problem-solving skills, in both experimental and applied contexts. This forms the unique contribution of the work, and helps delineate the underlying processes behind problem-solving abilities in WS and DS, in order to identify the constraining factors for performance. The opportunity to assess the relationships between experimental and real-life measures informs the generalisability of the experimental findings to individuals' daily lives.

In Chapter 2, problem solving and EF skills are measured experimentally in WS and DS and compared to a typical control group matched on nonverbal MA; the relationships between EF and TOL measures are also compared between WS and DS groups and a larger typical group, using a developmental trajectory approach. Chapter 3 presents a detailed exploration of TOL performance across the nonverbal MA-matched groups, both in terms of the behaviours exhibited during solving and how well individual EFs predict problem solving on different types of TOL trials. In Chapter 4, parental questionnaire data are reported from TD, DS and WS groups, examining both the patterns of performance on an EF questionnaire (the BRIEF) and a new questionnaire designed to assess everyday problem solving (the PSQ). In Chapter 5, the relationships between the experimental and questionnaire data are assessed. Chapter 6, the general discussion, forms a summary of the outcomes and conclusions, which are discussed in relation to existing literature, implications and potential future directions.

CHAPTER 2: EXPERIMENTAL MEASURES OF EXECUTIVE FUNCTIONS AND PROBLEM SOLVING

2.1 Introduction

Executive functions (EFs) can be considered as a set of related components in both children and adults (Lehto et al., 2003; Miyake et al., 2000) including planning, monitoring, inhibition, shifting and working memory. EFs are thought to facilitate goal-directed behaviour (Best & Miller, 2010). Problem-solving research addresses the way in which people organise their behaviour to work towards a goal: indeed, a fundamental characteristic of problem solving is the property of goal-directedness (e.g., Anderson, 2000, cited in Unterrainer & Owen, 2006). Problem solving on the Tower of London (TOL) task can be considered to be novel (Shallice, 1982) and complex (Berg et al., 2010). EFs are utilised when problems are novel, and require active monitoring and control (Rabbitt, 1997). Therefore, we would expect to see problem solvers drawing on EF skills in order to solve problems on the TOL task. Several studies to date (reviewed in Chapter 1) have investigated the relationships between EF tasks and TOL performance in typically developing (TD) children and adults. More recently, batteries of EF tasks have been used in research with individuals with Williams syndrome (WS) and Down syndrome (DS) in order to investigate EF profiles (e.g., Lanfranchi et al., 2010; Menghini et al., 2010). Most studies have measured EFs in a single disorder, rather than comparing performance in WS and DS, although there are some exceptions (e.g., Carney et al., 2013). Studies aiming to investigate and compare the contribution of EF skills to TOL problem solving in both WS and DS have hitherto not been conducted. The main aim of the current chapter is to assess the relationship between EFs and TOL performance in WS and DS using a cross-syndrome comparison.

2.1.1 Links between executive functioning and TOL performance

The potential links between EF abilities and TOL problem solving are reviewed in Chapter 1. Briefly, for TD children, there seems to be mounting evidence for the contribution of inhibition abilities to TOL performance (Asato et al., 2006; Baughman & Cooper, 2007; Bull et al., 2004; Kaller et al., 2008). Shifting/cognitive flexibility is another potential candidate (Anderson et al., 1996; Bull et al., 2004). Suggestions that working memory might be important but not adequately tapped into (Bull et al., 2004; Kaller et al., 2008) also raise the possibility of executive working memory, rather than more passive STM storage, being relied upon. The contributions of EF abilities to TOL performance have been suggested to vary along with the particular demands of the TOL task: for example, with reference to adults solving the TOH puzzle, Miyake et al. (2000) note that without directions as to how to go about solving the task, people may employ a perceptual strategy (working to increase similarity between the current state and the goal state as they go along) which may draw on inhibition skills. In contrast, Klahr and Robinson (1981)'s adaptation of the TOH task required children to produce a complete verbal plan, in advance, of all the movements that would need to be made. In Shallice (1982)'s original TOL participants were also asked to plan their moves before making them. This is likely to require more planning ahead, although the role of preplanning for efficient TOL performance has been challenged (Phillips, 1999; Phillips et al., 2001). The contribution of EF skills to problem-solving abilities is under-studied in WS and DS (see Chapter 1) and is addressed in this chapter.

2.1.2 The current study

2.1.2.1 Selection of EF tasks

In the current study, several EF tasks were administered along with the TOL, in order to investigate which of them might prove important for problem solving in individuals with neurodevelopmental disorders. Tasks were developed with the aim of being sensitive to a wide range of ages and ability levels, and for floor and ceiling effects to be avoided; to be physically easy to manage; to be nonverbal in the main, because the TOL does not require participants to produce verbalisations in order to solve it. The tasks were chosen to reflect existing proposed divisions of EF in the literature (inhibition, shifting and working memory). A planning task was also included, in order to be able to assess an individual's ability to plan ahead. In response to Bull et al. (2004) and Kaller et al. (2008)'s suggestions that their memory tasks were not accounting for performance on the TOL because they were measuring STM rather than working memory, both forwards and backwards versions of memory span tasks were conducted.

2.1.2.2 Selection of the TOL paradigm

With respect to the wide variety of approaches taken to TOL administration reviewed in Chapter 1, the selection of an appropriate TOL task and problem set was an important consideration in the design of this study. While some standard forms of the TOL task exist for which normative data are included and/or reliability has been established, there were several reasons for developing a unique problem set. Some forms (Anderson et al., 1996; Krikorian et al., 1994) use the original set of 12 problems from Shallice (1982)'s study, which have been noted to lack a linear increase in difficulty (Anderson et al., 1996) as well as having a maximum difficulty level of 5 move problems, and comprising a mixture of tower-, flat- and partial tower-ending problems. The latter point also applies to the version developed by Culbertson and Zillmer (1998). An additional version, (Schnirman, Welsh, & Retzlaff, 1998) developed for college students, did thus not guarantee suitability for populations with intellectual disabilities. One study has adapted the TOL for individuals with intellectual disabilities (Masson, Dagnan, & Evans, 2010) although they also used a problem set which included various types of goal configurations.

In the current study, the task was based on the standard design, referred to as the "Shallice TOL" by Berg and Byrd (2002) to allow comparisons to existing literature which also uses this design. Luciana and Nelson (1998) also noted that around half of the four-year-old children in their sample were unable to understand the SOC version, which is another reason to use a standard version. The task was developed to be appropriate for a wide range of participants, from four-year-old TD children and adults with developmental disabilities, to TD children 11 years of age. Thus, problems ranged in difficulty from a minimum of one move to a maximum of six moves. Huizinga et al. (2006) used problems with up to six minimum moves and reported that performance continued to develop on the task to at least age 15, which is beyond the age of the oldest TD participants in the current study. Masson and colleagues also used a range of problems from one to six moves, and found the task to be appropriate for their learning disabled population. A linear increase in difficulty was desirable, along with a threshold procedure, to allow participants to attempt problems which were appropriate for their ability level and to maintain participants' motivation.

Berg and Byrd (2002) give six colour permutations of six unique configurations of pieces (e.g. all pieces on top of one another, one on each peg, and four arrangements of partial tower configurations) and numbered them with a particular notation. This notation was utilised by Kaller et al. (2011) in TowerTool 2.0 open source software. Together, the software and the notation were employed in the current study to assess the parameters for various potential problems that could be selected. Two of the problem set used in the study were taken from Kaller et al. (2011)'s suggested set of problems. Newman and Pittman (2007) noted that goal hierarchy and the number of optimal solution paths available are important parameters to consider when choosing a problem set. Only problems with partial tower beginning and ending positions were chosen, to control for the goal configuration variable. Other parameters were also controlled for (optimal paths to solution) and others systematically varied (search depth; counterintuitive moves). See Appendix A for details of the problem set. Berg and Byrd (2002) recommended using multiple ways to measure performance rather than relying on a single indicator of success, including speed as well as accuracy. A variety of measures were thus included in the current study to gain an understanding of problem-solving performance: see Section 2.3.1 for details.

2.1.2.3 Aims and predictions

The aim of this study was to investigate problem-solving abilities in both typical and atypical development, and to understand how specific EF skills are used when solving problems. Regarding the relative EF profiles in WS and DS, it is expected that working memory tasks will reflect the existing verbal and visuospatial STM profiles found using span tasks, which is the type of task used to assess working memory in the current study: so, we can expect better verbal than visuospatial performance in WS, and the opposite pattern in DS. Given that tasks are predominantly nonverbal, we might expect this to be reflected in the EF profiles of the two groups – for example, visuospatial inhibition was poor in the WS group but not the DS group in Carney and colleagues' study. However, mixed results regarding EF profiles make predictions difficult: when WS, DS and TD groups were matched on nonverbal ability by Costanzo and colleagues, there were no group differences on visuospatial inhibition. In general, we can expect the contribution of EFs to vary with the difficulty of the problem that is presented, cf. e.g., Bull et al. (2004). There is evidence for inhibition contributing to TOL scores in young children, and improvements on the TOL with age, so this is also expected for the TD group. Based on Bull and colleagues' and Kaller and colleagues' suggestions that working memory might be more called upon than simply STM, this predicts that backwards memory tasks would be related more strongly to TOL performance than forwards memory tasks. There are some suggestions of shifting being related to the TOL (e.g., Bull et al., 2004), so we might also expect to see this pattern in the current study. If individuals are calling on planning ability to help them solve TOL problems, planning score should be related to TOL score. In the atypical groups, given the different EF profiles, we can expect each group to demonstrate different contributions of EF skills to TOL problem solving.

Comparing performance across typical and atypical groups should yield an understanding of the relative EF skills of each group: that is, whether the DS and WS groups' EF is impaired, and whether the EF profile of impairments differs between the groups. Comparing patterns of relationships between EF and TOL performance across syndromes can identify any syndrome-general or syndrome-specific patterns of associations. This is an important area of investigation for several reasons: if we know which skills people draw on when approaching a problem, this would help to point the way towards appropriate interventions. Identification of differences in the relative contributions of EF skills between typical and atypical groups might highlight potential compensatory strategies. Understanding how compensatory strategies might be utilised by individuals with neurodevelopmental disorders would further inform how intervention studies could be approached. Finally, including a battery of EF tests for both disorders allows us to add to the existing, somewhat limited, literature investigating EF skills in these two populations.

2.2 Method

This study comprised the administration of several experimental tasks. A consideration of the participants and overarching design and procedure will be followed by details of the materials and procedure for each individual task.

2.2.1 Participants

In total, 96 participants took part in the study, comprising 20 individuals with WS, 20 with DS and 56 TD children. For the WS group, only individuals who had received a positive phenotypic diagnosis and genetic confirmation of the disorder, in the form of a fluorescence *in situ* hybridisation (FISH) test, were tested. Exclusion criteria for the TD group were: known premature birth, special educational needs or medical conditions which may have affected their ability to complete the tasks. One participant was found to have a recent diagnosis of colour blindness and was excluded. One participant with WS was unable to understand several tasks and was also excluded. Of the 56 TD participants, a subset of 20 was individually matched by a measure of nonverbal ability, the Raven's Progressive Coloured Matrices (RCPM; Raven, 2004) to the DS and WS groups. One individual from each group was matched individually to one participant from each of the other groups, resulting in three individually-matched groups consisting of 20 'trios' of RCPM-matched participants. For the vast majority of trios the RCPM score differed by no more than 3 points in total (across all three participants), while one trio differed by 4 points and another by 10 points. RCPM score was equivalent across the three matched groups ($F<1$). Receptive vocabulary was also measured using the British Picture Vocabulary Scale, third edition (BPVS-III; Dunn et al., 2009) which differed significantly across the three matched groups ($F(2,57) = 16.789$, $p < .001$, partial $\eta^2 = .371$), with Tukey *post-hoc* tests revealing better scores in the WS than the DS or TD groups ($p < .001$ for both) and equivalent scores between the DS and TD groups ($p = 1.0$), reflecting the expected elevated receptive vocabulary scores for a WS group. Participant demographics are displayed in Table 2.1.

	Males: Females	Total N	CA Range (years: months)	CA (months)	RCPM Raw Score	BPVS Raw Score
				Mean (SD)		
DS	11:9	20	12:7 – 24:2	207.25 (41.49)	17.05 (5.52)	83.75 (24.90)
WS	11:9	20	12:4 – 24:3	212.65 (43.15)	17.75 (3.77)	117.35 (22.22)
TD	28:28	56	4:4 – 11:5	94.20 (24.97)	25.79 (7.47)	106.98 (24.44)
TD Matched	11:9	20	4:4 – 10:2	69.85 (16.00)	17.65 (4.25)	83.85 (15.07)

Table 2.1: Participant demographics (experimental work)

2.2.2 Ethical considerations

The project was granted ethical approval and informed written consent was obtained from all participants' parent or guardian. All participants gave their verbal assent to take part, and in addition all participants of 12 years and above gave their written consent after an explanation of the study. Care was taken to inform participants of the anonymous nature of the video recordings. Participation was voluntary, breaks were given as necessary during testing sessions, and participants were free to terminate a session if they needed to. A range of motivation tools was used, including positive praise and encouragement and a sticker added to a participant's named card on completion of each task. All participants were offered a sheet of stickers as a reward for taking part in the study. When visiting the university, travel expenses were reimbursed. No prior incentives were given for participation.

2.2.2.1 Recruitment

The TD children were recruited through and tested at their primary school in London. The participants with WS were recruited through the Williams Syndrome Foundation: parents who had previously agreed to be contacted were sent an invitation for their son or daughter to take part. The experimenter generally visited the participants with WS at their homes to conduct the research: one accompanied participant visited the university. The participants with DS were recruited from existing links with families from within the research group. The experimenter visited participants at their homes or at their schools.

2.2.3 Overall design and procedure

For each task, the independent variable of Group had three levels (WS; DS; TD). All participants completed the Tower of London, five executive function tasks measuring shifting, planning, inhibition, verbal working memory and visuospatial working memory, as well as the BPVS and RCPM mental age measures. Testing took a total of approximately two hours for each participant, with breaks, with durations varying dependant on individuals' pace and ability. Several factors were taken into account to determine the number of sessions necessary, including the rare nature of WS and thus the long distances necessary to travel, the limited attentional capacities of young children and individuals with neurodevelopmental disorders and the practical aspects of conducting testing sessions within normal school hours. Thus, testing took place over one (for participants with WS), two (for participants with DS) or three (for TD participants) separate sessions. For all sessions breaks were included as needed. The order of tasks was counterbalanced.

Participants were tested individually in a quiet room or separate area and completed the tasks seated at a table with the experimenter. Each task was explained separately and participants were praised throughout for their efforts and achievements. Where possible, tests were concluded by requiring the participant to answer a part of the task with which they had experienced success. Verbal encouragement and praise was given as appropriate, at intervals throughout all of the tasks.

2.2.4 Task-specific outlines

2.2.4.1 Mental Age (MA) measures

Tasks were administered in accordance with the published instructions.

2.2.4.1.1 BPVS

The BPVS-III (Dunn et al., 2009) is standardised for ages 3 to 16 years. Participants are repeatedly shown four coloured pictures on a page and asked to choose the picture that goes best with an auditorily presented word, spoken by the experimenter; in this way an index of receptive vocabulary was ascertained, in the form of the raw score. The task took between 10 and 20 minutes to complete.

2.2.4.1.2 RCPM

The RCPM (Raven, 2004) is standardised for ages 4 to 11 years. Participants are shown 36 patterns with a section missing and asked to select the correct piece, from a choice of six, to complete the pattern. All items are administered for each participant. The task took approximately 5-10 minutes to complete.

2.2.4.2 The Tower of London (TOL)

2.2.4.3 Design

Following a demonstration trial and two practice trials, participants completed several experimental TOL trials. Four 2-move trials were presented first, provided that the participant was able to complete the two 2-move practice trials. Trials proceeded with two of each of 3-, 4-, 5- and 6-move problems until the end of the problem set or until a participant completed both of the trials of one length incorrectly. If fewer than half of the 2-move trials were completed correctly, four 1-move trials were also administered. Thus, while there was a total of 16 experimental trials, in practice a maximum of 12 experimental trials was completed (all of the 2-6 move problems). The full problem set is displayed in Appendix A. The principal dependent variable was the TOL Score (see Section 2.3.2) while several other dependent variables were collected from the video recording of the task (see Section 2.3.1).

2.2.4.3.1 Apparatus and materials

The TOL apparatus is represented in Figure 2.1.

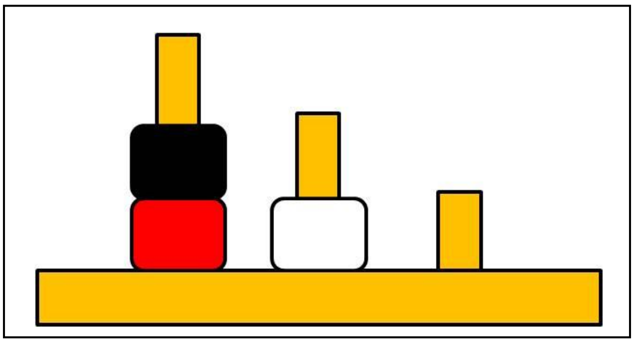

Figure 2.1: Schematic representation of Tower of London (TOL) apparatus

The apparatus comprised a wooden board and additional pieces, as shown in Figure 2.1, with approximate measurements as follows: a base of 24cm by 8cm by 1.7cm, with three wooden upright pegs spaced equally along the board at 6cm intervals. The pegs were approximately 2cm in diameter and were of varying heights: the left hand peg protruded 7.5cm above the base, the central peg 5cm and the right hand peg 2.5cm above the base. Three ring-shaped pieces (red, white, black) were provided, measuring 4.2cm at their widest point and 2.4cm high, which fitted comfortably onto the pegs, such that the left, centre and right pegs would hold three, two and one pieces respectively. The card with the goal state printed on it was rested on a blank stand so it could be viewed easily by the participant, and was stood between the stand and a transparent plastic cylinder attached to the table horizontally to keep it in place. Goal states were in the form of a schematic diagram of the game board (as in Figure 2.1), printed onto card and laminated. When not in use (between trials) pieces were placed on a piece of card the same colour as the wooden board. The task was recorded on a Sanyo Xacti SD card video recorder (VPC-65EX and VPC – HD2EX) supported by a tripod.

2.2.4.3.2 Procedure

Participants were seated at a table with the apparatus placed at a comfortable distance. The video recorder was switched on and the experimenter introduced the task. Participants were shown the three pegs and their different sizes were pointed out. Participants were told that three pieces would fit on the big peg, two on the middle and only one on the small peg. They were asked to identify the colours of the pieces, and then shown a goal state picture and asked to show the experimenter where each piece was on the goal state, in order to establish that all participants could distinguish and identify the different coloured pieces. At the start of each trial the card was turned over so that the goal state was visible. A 1-move demonstration trial began the task procedure. During the demonstration trial the participant was instructed that they would be moving the pieces so that they matched the picture. The rules were explained and demonstrated: that only one piece could be moved at a time, that the pieces needed to stay on the pegs and could not be placed, for example, on the table, and that no more pieces could be balanced on top of a peg if it was already full. Participants were informed that they needed to try and solve the problems in as few moves as possible. Then the experimenter completed the 1-move demonstration trial, pointing out the move (e.g., "I just need to move this one here") so that the solution was correct. The experimenter then asked the participant whether the game board matched the picture, allowing her to ascertain that the participant could judge that the two arrangements of pieces were the same or give further explanations if necessary. Two 2-move practice trials followed the demonstration, to allow participants to move the pieces in accordance with the rules and make their pieces match the goal state. If a rule was violated during a practice trial, the experimenter praised the participant's effort and showed them where the rule had been broken, and restarted the practice trial from the beginning. Practice trials were repeated until the experimenter was satisfied that the participant had understood the task.

Experimental trials proceeded in a sequential fashion. At the beginning of each trial the experimenter set up the pieces in the starting configuration and placed the goal state facing away from the participant. She said, "Ready? Off you go", and turned the goal state over so that it was visible to the participant. The goal state remained visible throughout the trial. Participants moved the pieces until they were satisfied that they had matched the picture. The number of moves made was recorded during the trial by the experimenter. Where a rule was violated, the experimenter stopped the participant, reminded them of the rule and pointed it out (e.g., "stop there – you've got the red and the white pieces moving together there [holding up the pieces]. Keep going, but remember that it's just one piece at a time"). The experimenter then returned the pieces to the positions that they occupied immediately prior to the rule violation, and the participant was asked to continue. A trial ended when a participant said that they had finished. Alternatively, the experimenter ended the trial when the participant had made more than 20 moves and was still unable to solve the trial ("that's a bit of a tricky one, isn't it? You've had a really good go. Let's go on to the next one") or a participant ended the trial by reporting that they were unable to complete it. Trials proceeded according to the threshold procedure described above. A trial was judged to be correctly completed when the goal state was matched in 20 moves or fewer. Participants were not permitted to restart trials: occasionally the experimenter restarted a trial where it had not been possible to correctly replace the pieces after a rule violation. The task took between 10 and approximately 30 minutes to administer, depending on the participant's individual pace and ability.

2.2.4.4 Executive function tasks

2.2.4.4.1 Inhibition Task

In this task, participants were repeatedly presented with an image on a computer screen and asked to respond by pressing one of two pictures that were attached to the keyboard.

Design and procedure

A repeated-measures design was employed such that all participants completed both a 'same' condition and an 'opposite' condition. The same condition was always administered first, and the two conditions were separated by the planning task (see below). Two versions of each condition were created such that half the participants completed the same condition with images of grass and snow, while the other half were given leaves and stones. Each participant subsequently completed their opposite condition using the other set of pictures. The dependent variables were accuracy and reaction time (RT) to each condition as well as the proportional percentage difference in the number of correctly solved trials and the proportional difference in the average time taken to complete correct trials, between the two conditions. Each condition consisted of 16 experimental trials and took around 2-3 minutes to complete. A schematic representation of the task can be found in Figure 2.2.

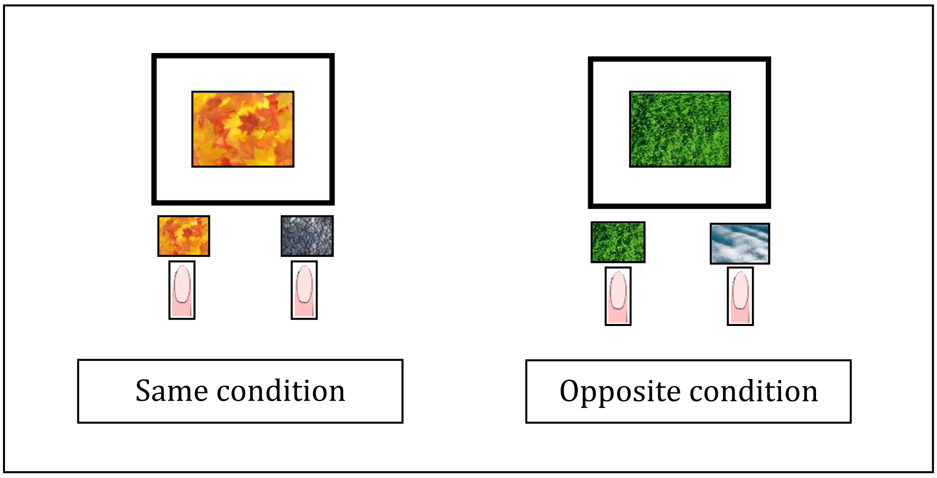

Figure 2.2: Schematic representation of inhibition task. Half of the participants saw Leaves/Snow images in the Same condition and Grass/Snow in the Opposite condition while the other half saw the reverse pattern.

Same condition

The task began with a welcome screen displaying the word 'Hello' in the centre of the screen. The experimenter consecutively displayed the two images for the condition, explaining that the participant would see the picture of (grass/leaves) or the picture of (snow/stones). The experimenter pointed out the pictures on the keyboard as well, naming and pointing to them. Participants were told that they should press the picture on the keyboard that was the same as the one on the screen. In a practice phase, four trials followed in which participants demonstrated their understanding; three out of four were required to be matched correctly before continuing on to the experimental phase. If this did not occur, the practice trials would be repeated and further explanation given.

Participants were then introduced to a sound that would play when they responded correctly, and told to keep going if they made a mistake. Speed and accuracy were both emphasised and a final reminder given of the rule (i.e., pressing the picture that is the same). Participants were asked to hold their hands just above the pictures so that they were ready to press them. Each image was displayed until a response was made. A correct response produced the short 'twinkly' sound from the computer and the image was followed by a blank white mask and an 850 msec pause before the next image was presented. An incorrect response produced no sound and no change to the image. When incorrect, the experimenter reminded the participant of the relevant rule. When a subsequent correct response was made, the task proceeded as for a correct answer. The task ended with a 'well done' screen, with a large star in the centre of the screen, a smiley face and balloons, irrespective of performance.

Opposite condition

The Opposite condition proceeded in exactly the same way as the Same condition, with participants instructed to press the picture that was different from the one on the screen. Sometimes the term 'different' was not understood and substituted expressions were used, e.g. the one that's not the same, or the other picture, with practice trial blocks being repeated until the participant had understood.

Apparatus and materials

Images of grass, snow, leaves or stones, measuring 15cm by 10cm, were presented in the centre of an Acer Aspire 5810TZ widescreen laptop measuring 15.6 inches using Superlab 2.0 software. Two laminated pictures of the two relevant images were attached to the computer keyboard, each measuring 4.2cm by 3cm with a small white border. They covered keys *(A S D Z X)* for the left-hand image, and keys *(K L ; , .)* for the right-hand image. The pictures of grass and leaves were always attached to the left-hand side and the snow and stones to the right, according to the relevant version.

2.2.4.4.2 Planning task

Design and procedure

The planning task was designed to require participants to plan several steps ahead. Like Fagot and Gauvain (1997)'s and Carlson, Moses, and Claxton (2004)'s delivery tasks requiring items to be placed in reverse order, the planning task here was also designed to tap into this type of ability. Participants loaded small wooden 'boxes of milk' onto a milk van for delivery to houses. To succeed with the task, milk boxes needed to be loaded in reverse order. The trials increased in difficulty by increasing the number of houses that required milk. The dependent variable was the number of correct trials. The task consisted of three phases in total. Testing was terminated when two consecutive trials were completed incorrectly within any one phase.

Phase one

Participants were comfortably seated at the table. The experimenter showed the participant the board and explained that they were going to pretend that they had a job in a town, helping to deliver some milk. The houses were placed in a row in front of the participant and they were asked to identify the colours, (either by saying the colour names or pointing to the colour name said by the experimenter) in order to check that they were able to distinguish them. All of the boxes of milk were then placed in the centre of the road, and it was pointed out that the boxes of milk were different colours as well, and explained that the boxes should go to the house of the same colour. To check understanding, the experimenter demonstrated that the red box would go to the red house, and asked the participant where the blue box would go. Where any difficulty with naming or identifying colours was shown, these colours were also selected for the box-house matching to check that participants were able to match the boxes to the houses. All participants were able to do this (even if they did not know the names of the colours), with the exception of one who was excluded from the study (see Section 2.3.4.2). The houses were then lined up to one side of the road.

The van was shown to the participant and placed in the centre of the road, and the plastic tube inserted into the top, while the experimenter explained that its purpose was to facilitate the placing of items onto the van. The starting point (flag) and arrows on the road were pointed out to the participant and it was made clear that, “as the arrows only go this way round, it’s a one way street, so the van can only drive this way around”.

A demonstration trial followed in which one (red) house was placed on the board. The experimenter showed the participant how the milk boxes were placed onto the van (by threading the box onto the tube) and demonstrated this while the van was in the centre of the road, at the loading bay. Then the van was driven around the track and the experimenter said, "I can give the red box to the red house", and did so. The van was driven back to the starting point and the participant was informed that "that one's finished". Another demonstration trial followed, with the experimenter explaining that, e.g., "when the van gets to a house, it can only deliver the box that's at the top [pointing to the top of the tube]. So when you're putting things on the van, you need to be careful about the order that you put them on the van". The experimenter loaded the two boxes of milk onto the van, placed it in the start position and began driving. "Can you see that I've made it so that the orange one is at the top, so that it can go to the orange house? And then [driving to next house] I can give the green one to the green house". A 2-house practice trial followed in which the participant loaded the milk onto the van for the experimenter to drive around the track. It was repeated with further support and explanation until successfully achieved.

The experimental trials then began. The participant's task was to load the boxes of milk onto the van for the experimenter to drive around the track. Visual feedback was clear throughout the task as the participant could easily see when they had loaded the correct or incorrect colour. The van was driven clockwise around the track, stopping at each house and delivering it if correct, with a "here's this one… well done…" etc. If the van reached a house with a non-matching (incorrect) colour at the top, the experimenter pointed out the mistake and placed the box on the loading bay. The participant was reassured, but reminded of the rule (to remember that it is the one at the top that can be delivered to the same coloured house). After each trial, all of the milk boxes were returned to the centre of the road and all the houses to the side. The experimenter then set up the next set of houses, adding them to the board in a clockwise order. When loading boxes of milk onto the van, sometimes a participant would remove some items from the van and replace them with alternatives. This was only allowed before delivery of the milk to the houses had commenced. There was one experimental trial of delivery length (number of houses) of two, three, four and five houses. A maximum number of replacements was set at twice the number of houses to be delivered to. If this maximum number was exceeded (for example if a participant continually removed and replaced items) the trial was terminated and scored as incorrect. This occurred very rarely; most participants presented a set of items to be delivered.

Phase two

In phase two, participants were informed that they were going to look at a different town where things happen differently. The phase one board was removed and replaced with the phase two board to reinforce the idea that it was different (in fact the phase two board only differed in the placement of background trees; see Figure 2.5). Participants were informed that the milkman in this town still delivers milk, but that sometimes there was a birthday party at one of the houses. Laminated cardboard balloons were attached to four of the houses (red, green, yellow and black), and the corresponding four boxes of milk were removed from the participant's view. It was explained that when there was a party, there were balloons outside the house, and the milkman would be delivering a birthday cake. Four birthday cakes made from felt and sponge (see Figure 2.4) were shown to the participant. Great care was taken to construct the cakes so that they would hold their shape, but would become squashed when a wooden block was placed on top of them. The participant was asked to identify the colours of the cakes as for phase one, and told that the cakes go to the house that is the same colour, and asked which house the yellow cake would go to in order to check understanding. All participants were able to pass this check. Participants were told, "we have cakes and milk to deliver now. But, watch what happens when I try to put milk on top of a cake". A cake was put onto the van, and a box of milk dropped on top so that it squashed the cake. The participant was asked what had happened. If they did not respond the experimenter said, "it's got squashed, hasn't it?". All participants were able to appreciate this. The experimenter continued, "No one wants a squashed birthday cake! So when you put the things on to the van you need to make sure that the milk is at the bottom, and the cakes are at the top [pointing to relevant part of post] so that the cakes don't get squashed". Participants were told that the van would now be driven around the road twice: once to deliver the cake, and again to deliver the milk. A 2-house demonstration and practice trial followed to illustrate the changes to the procedure. The trials then proceeded as for phase one, with a 6- and 7-house trial also included. If on the very rare occasion that a box of milk was loaded on top of a cake (thereby squashing it) the participant usually noticed and corrected this themselves. If they did not, they were told, "oops! Don't squash the cake!" This was recorded and the pieces removed. The participant was allowed to begin the trial again with a reminder not to squash the cakes. The items were delivered by the experimenter, commenting "round we go for cakes… and round we go for milk…" etc.

Phase three

This phase proceeded as for phase two, except that the new town was presented with a tree across the road (see Figure 2.5). The participant was told that there had been a thunderstorm and asked if they could see what had happened. It was explained that as the tree was across the road, the van could not drive past it and would have to turn back when it got to the tree: "so the van goes around this way for cakes, but has to go back the other way to deliver the milk". Additional arrows were added to the road, with the experimenter explaining that although it was a one way street, the arrows were a reminder that the van would be driving in both directions.

Apparatus and materials

The planning task was presented on a table top and is displayed in Figure 2.3.

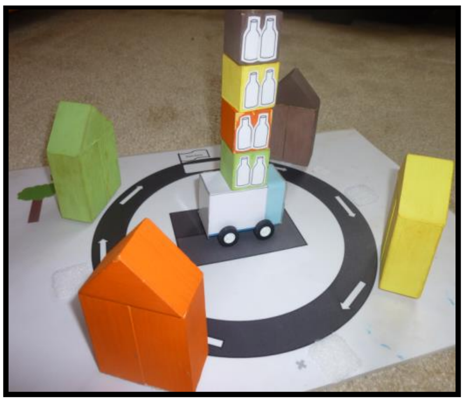

Figure 2.3: Planning task apparatus

An A3 matt-laminated board was attached to the table surface, featuring a circular road of 24.5cm diameter at its widest point, with a ‘loading bay’ in the centre and a flag at the top identifying where the van should start driving from. Houses were coloured and plastic-covered wooden blocks measuring approximately 6cm by 9cm by 3cm, and a small van measuring 6.5cm by 4cm by 3cm was placed in the centre. A fixed plastic tube protruded 23.5cm above the top of the van. Boxes of milk were small coloured wooden cubes of 3cm with a hole drilled through the centre of approximately 1cm, so that they could be comfortably threaded onto the tube. Pictures of milk bottles were affixed to two opposite faces of each cube. There were eight houses of various colours (red, blue, brown, black, yellow, green, purple, orange) and eight milk boxes of the same colours. Additional apparatus comprised four cardboard balloons and four birthday cakes made from felt and sponge, and two additional game boards for phases two and three. Figure 2.4 displays an example of a house with balloon attached, cake and milk box, while Figure 2.5 displays the phase two and three game boards.

Figure 2.4: Apparatus for planning task: house with balloon, cake and milk box

Figure 2.5: Game boards for phases two and three of the planning task

2.2.4.4.3 Shifting

Design and procedure

This task draws on elements of the dimensional change card sorting task (DCCS; see Frye, Zelazo, & Palfai, 1995; Kirkham, Cruess, & Diamond, 2003; Zelazo, 2006) and Wisconsin Card Sorting Task paradigm (WCST), while gradually increasing in difficulty in stages, as does the set shifting task from the Cambridge Neuropsychological Test Automated Battery (CANTAB; e.g. Sahakian & Owen, 1992). Participants were required to sort items by one dimension (shape or colour), and then make a shift and sort them by another dimension. Repeated shifts between dimensions were required. The dependent variable was the number of shifts that were successfully made.

The task was divided into three phases. Each phase needed to be passed in order for the next phase to be attempted. To pass a phase and progress to the next one, at least one shift needed to be made. To make a shift it was necessary to pass two consecutive blocks within a phase (i.e., with different sorting criteria). To pass a block, four consecutive trials needed to be completed correctly. The experimenter recorded the outcome of each trial as it occurred (correct or incorrect) to keep track of the score. Each block ended with a 'well done' screen (identical to the one displayed for the inhibition task) regardless of performance, the experimenter praised the participant's efforts, and the task was either ended or continued dependent on performance. The task took between approximately 10 and 25 minutes to administer, depending on performance.

Introduction and practice phase

The task began with a welcome screen with the word 'Hello' in the centre. Participants were shown a green triangle at the top of the screen, followed by two shapes appearing simultaneously at the bottom of the screen: a yellow triangle and a green circle. Participants were first shown the shape at the top, and then the two shapes at the bottom. It was explained that to play the game participants would need to determine which of the bottom two shapes the top shape belonged with. They were instructed that they could select a shape by tapping the screen. The participants were then played the 'twinkly' correct feedback sound, and reassured not to worry about any mistakes. They were then shown the first game ("this is the colour/shape game"). If a participant was first introduced to the colour game, they were subsequently introduced to the shape game, and the shape game was then the sorting criterion for the first block of their experimental trials. The reverse occurred when the shape game was the first to be introduced. Half of the participants completed each of these conditions.

They were asked to identify the green/yellow items (or the circle/triangle depending on the game) and told, "so, in the (colour/shape) game, the green (circles) ones go here and the yellow ones (triangles) go here". They practised the colour/shape game: four practice trials were given in which three were required to be correct to progress (see Figure 2.6). The practice trial block was repeated with extra explanation and support if this was not passed at first, until understanding was demonstrated. Then the alternate game (shape/colour) was introduced and practised in the same way. Examples of practice trial stimuli are displayed in Figure 2.6.

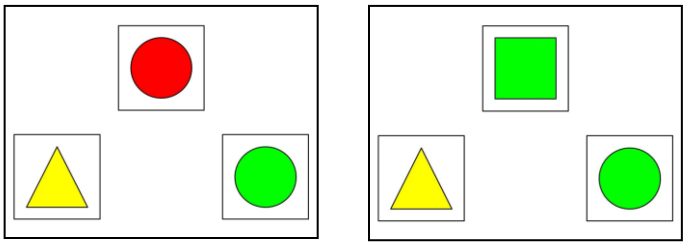

Figure 2.6: Initial practice trials in phase 1 for shape (left hand side) and colour (right hand side) criteria of the shifting task

Phase one

The first block of experimental trials in phase 1 then began: the experimenter informed participants that to start with they were going to keep on playing the shape/colour game (whichever game they had been introduced to second). They were asked to identify where the different types of images would need to go as a reminder. At the start of each block the experimenter introduced the sorting criterion: "now this time we're going to play the shape/colour game! Where are the green ones/yellow ones/triangle/circles (as appropriate) going to go?". During a block, if incorrect responses were given the experimenter would remind the participant of the game, e.g., "Oops! Remember it's the shape game". Throughout the task, incorrect answers did not elicit a response from the computer: the participant was given the opportunity to self-correct (and thus gain the correct feedback sound) before the next trial appeared.

Four blocks of eight experimental trials were presented, either requiring sorting by colour ('the colour game') or shape ('the shape game'). The shape to be sorted (at the top) was of the form of one of the two displays in Figure 2.7, presented in a fixed pseudorandom order of trials, and the shapes at the bottom were always a yellow triangle and a green circle. In this way it was always possible to match the stimulus to either of the shapes at the bottom; the top shape always corresponded to one of the bottom shapes by one of the dimensions. The answers which were correct or incorrect varied with the current sorting criterion (colour or shape). Examples of stimuli in phase 1 and 2 are displayed in Figure 2.7.

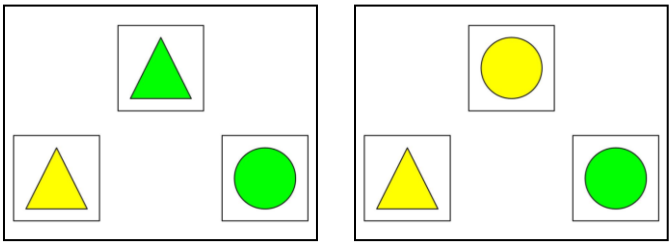

Figure 2.7: Experimental stimuli in phases 1 and 2 of the shifting task

Phase two

Phase two was identical to phase one with the exception of the new requirement to use the feedback (a 'twinkly' sound for correct responses only) provided by the computer to discover and follow the rule; now sorting criteria were not told to the participant. The first trial in each block was not counted towards the four consecutive correct trials necessary to pass, because one ('elimination') trial was needed in order to determine the rule. Participants were informed that they would be doing the same as before except that they would not be told the rule and needed to listen to the sounds in order to work out which game they were playing. They were informed that "sometimes the computer will change the game that it's playing, so you need to keep on listening and see if you can figure it out". At the start of the phase they were asked where the images would need to go, as a reminder (e.g., "if it's the shape game, you're going to put triangles…? And circles…? If it's the colour game, you're going to put green ones…? And yellow ones…?"). In this phase the four blocks with different sorting criteria were presented (the rule still changed every eight trials) but there was no pause or input between blocks. Encouragement was given throughout. If a participant seemed to be getting many trials incorrect the experimenter encouraged them to try to work out the game (e.g., "is it the colour game or the shape game?").

Phase three

Phase three proceeded as for phase two, but another dimension (number) was introduced. The top shape now corresponded to the bottom shapes on all three dimensions (see right hand side of Figure 2.8 below). Four blocks of 12 trials were presented, for which the rule (colour, shape or number) had to be discovered and followed. For participants who reached phase 3, the 'number game' was introduced and four practice trials administered (see Figure 2.8 for an example). Participants were given instructions regarding listening to sounds and working out the game in the same way as for phase two. Because the shape at the top of the screen corresponded to the ones at the bottom of the screen in three ways (shape, colour and number), the first trial in each block was carefully selected so that after the first ('elimination') trial, it was possible to discover the rule on the next trial. Thus, four consecutive correct trials, not including trial 1, constituted passing a block, as for phase two.

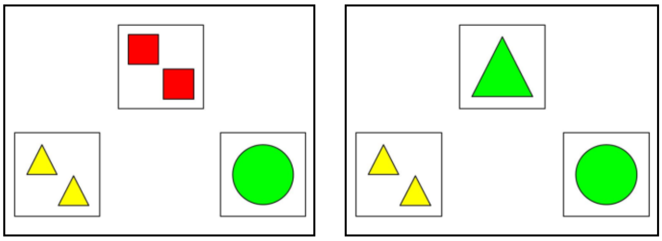

Figure 2.8: Example stimuli from phase 3 of the shifting task: training (left hand side) and experimental (right hand side)

Apparatus and materials

The task was presented on a touch screen computer (Toshiba Portégé M780-112) with a 12.1-inch screen. Example stimuli are displayed in Figure 2.6, Figure 2.7 and Figure 2.8 above.

2.2.4.4.4 Working memory

Verbal and visuospatial working memory tasks were administered, involving repeating sequences of digits or tapping sequences of blocks, respectively. Each task had a forwards version and a backwards version, and the scores were in the form of a span measure: the longest list length that could be successfully remembered. The backwards tasks were administered immediately after the forwards tasks (in line with Wechsler & Naglieri, 2006).

Visuospatial working memory

Design and procedure

The task was based on the spatial span task from the Wechsler nonverbal scale of ability (WNV; Wechsler & Naglieri, 2006), with a modified procedure, outlined below. Children were shown the board and told that in this game they were going to be tapping some blocks. The experimenter explained that they would tap a block and the participant should tap the same one. Practice trials for 1 item then ensued. The participant was then told that this time the experimenter would tap more than one, and that they should try to tap the same ones, in the same order. Two 2-item practice trials then followed, with repetitions, further explanations and demonstrations of the correct responses where appropriate until the participant understood the task. The experimental phase consisted of two trials of each length, from 2 up to 9 items long. Testing continued until both trials of a particular sequence length were reproduced incorrectly.

In the backwards version of the task, participants were told that the game they would be playing was now different. Practice trials began with 1-item trials and then participants were told that this time when the experimenter tapped more than one, they should try to tap the same ones, but backwards. Practice trials of 2 items in length followed, then experimental trials of 2 items in length, increasing to 3, 4, 5, 6, 7, 8 and 9 until the participant incorrectly reproduced both trials of the same length. Pickering and Gathercole (2001) noted that children may need special direction on how to reverse three items. In a departure from Wechsler and Naglieri (2006)'s procedure, where any participants were correctly able to reverse a sequence of two but not three items, additional demonstration and practice trials were given in reversing three items, followed by experimental trials of three items, to allow participants to demonstrate whether or not they were able to reverse three items once they had been instructed in how to do this (cf. Pickering and Gathercole (2001)'s instructions: the last one, then the middle and then the first). Five out of the 20 TD participants in the matched group demonstrated the need for this support version; as did 10 of the participants with WS and 11 of the participants with DS. Also, where a participant reproduced a 3-

item trial sequence in the same order as the experimenter, this was pointed out and the trial repeated with a reminder that the order should be reversed. The task then continued as before until both items of the same length were completed incorrectly, or until the end of the trials.

Apparatus and materials

The spatial span board, provided with the WNV battery, consisted of a white plastic board with ten blue cubes protruding from the base (see Figure 2.9). Printed on the experimenter's side of the cubes were the numbers 1-10 for identification.

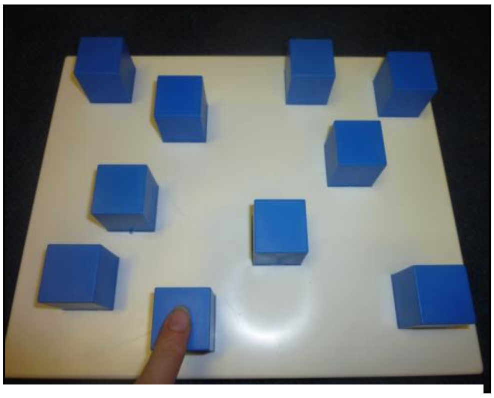

Figure 2.9: Visuospatial memory task apparatus

Forwards verbal working memory

Design and procedure

The Digit Recall sub-test from the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) was used. Two participants responded in a modified way as they were unable to produce an audible response (see the Apparatus and materials section below). The task was administered in accordance with the instruction manual, with a modified structure (in order to maintain consistency with the visuospatial working memory tasks) which is outlined below. Participants were told that the experimenter would be saying some numbers and that they should listen to the numbers and say the same ones back; to copy what the experimenter was saying. There were two practice trials of one number followed by two practice trials of two numbers. Practice trials were repeated with further explanations until the task was understood, with answers to practice trials given if necessary.

The experimental trials involved the experimenter first informing the participant of the number of numbers that she would say, in line with the WMTB-C test instructions. Numbers were said aloud at a rate of approximately 1 number per second and participants asked to repeat the list of numbers. Participants were given encouragement during the task and told when the number of numbers to be repeated would change. Two trials of each length were administered, with the task being terminated when both trials of the same length were reproduced incorrectly. These trials were the first two trials of each block from the WMTB-C trial set. As for this task, trials up to a list length of 9 digits were included in the set: that is, the maximum number of trials administered was 16 (two each of 2-, 3-, 4-, 5-, 6-, 7-, 8- and 9-item trials). If a participant asked for the string to be repeated, this was done but no credit was given for the trial. Occasionally the trial was repeated correctly but with additional items produced, and these were scored as incorrect.

Apparatus and materials

For the majority of participants the forwards condition did not require any apparatus or materials. For two of the participants with DS who had difficulty producing numbers verbally, number cards from 1 to 9 were produced by the experimenter for the participants to point to in lieu of vocally producing the numbers.

Backwards verbal working memory

Design and procedure

The backwards version of the verbal working memory task was modified from the Backward Digit Recall sub-test from the WMTB-C, in the same way as the forwards version. One participant with DS was unable to understand the instructions and so did not provide data for this task. Where necessary, counting forwards and then backwards using the number stimulus page was practised with the participant to ensure understanding. Practice trials followed in which two number cards were shown to the participant with the experimenter saying the numbers forwards and asking the participant to say them backwards, and then two more practice trials in which the participant watched the experimenter point to two blank cards in order and say a number for each, and was asked to repeat the numbers backwards. The experimental trials then proceeded in the same way as for the forwards verbal working memory task, with participants asked to repeat items in reverse, except for the modification that is suggested in the WMTB-C manual: that is, after the 2-item experimental trials were completed successfully, practice trials were administered for 3 items. As Pickering and Gathercole (2001) note, this is because being able to reverse 3 numbers is a skill in itself, and a participant may fail on 3 item trials because of a lack of this understanding, rather than memory limitations. Reminders about the backwards rule were provided as needed. As in the backwards version of the block span task, where a participant responded to a trial with the correct numbers in the same order as the experimenter, this was pointed out and the participant given further attempts at producing the numbers in the reverse order.

Apparatus and materials

The stimulus page from the WMTB-C manual (with the numbers 1-9 and forwards and backwards arrows) was used to help demonstrate counting backwards for all participants, with the exception of the TD children who did not require it (who were usually at the older end of the age range). Supplied number cards were used for all participants in the demonstration and practice phases. For two of the participants with DS who had difficulty producing numbers verbally, number cards from 1 to 9 were produced by the experimenter for the participants to point to in lieu of vocally producing the numbers.

2.3 Results

2.3.1 TOL coding

TOL performance was coded from video recordings of the task. A summary table of the measures coded are displayed in Table 2.2, along with further descriptions of task behaviours relevant to the analysis. Some measures (e.g., rule violations, type of error) will be discussed in Chapter 3, in which more qualitative aspects of performance are explored. The measures in **bold type** contribute towards the TOL score.

Name	Description
Move (or peg-to-peg move)	The movement of a piece to another peg by the participant. To be classed as a move, the piece must leave the original peg and at least touch the top of the new peg
Incomplete move (1 move point)	A piece is lifted off its current peg, then held above the pegs and returned to its original position
Backup move pair (2 move points)	Where two moves are made (whether the piece is let go of or not) and the piece is returned to the peg that it has just left
Hover (half a move point)	Approaching a peg with a piece then hovering above it without physically placing it onto the peg
Rule violation (1 move point)	Lifting two pieces off the pegs at once, placing a piece elsewhere than on a peg (e.g., on the table), or balancing a new piece on top of a peg that is already full. Rule violations that are immediately self-corrected are not counted
Verbalisation	A binary judgement for each trial of whether a relevant verbalisation is made to help solve the trial, for example saying 'the red goes there…then the black there…'
Planning time	The time elapsing between the turning over of the goal state (goal presentation) and when the participant makes the first move of the trial
Total trial time	The time elapsing from the goal presentation (when the goal state card has been turned over and the bottom edge reaches the surface) to the placement of the last piece that constitutes the final move
Execution time	Total trial time – planning time
Time per move	Execution time / number of moves made

Reason for ending trial	Whether the trial is reported as finished by the participant, whether more than 20 moves are made and the trial is ended by the experimenter or whether the participant ends the trial because they are unable to solve it.
Error type: perceptual	Trials reported as finished where the configuration of the pieces does not match the goal state

Table 2.2: Descriptions of TOL coding measures and behaviours

Recordings for a random 10% of participants were also second-coded. Inter-rater reliability was coded for the number of moves made (up to 20), incomplete moves, backup moves, whether a trial was completed correctly in 20 moves or not, rule violations, hovers, verbalisations and the reason for ending the trial, and disagreements settled by discussion. Spearman's correlation coefficient was used to calculate inter-rater consistency, and the lowest correlation coefficient was r = .847 (p < .001) (for verbalisations), indicating a high level of agreement for each variable, as an r value of .7 is deemed acceptable (Multon, 2010). In addition, the percentage agreement (percentage of identical coding judgements) was above 90% for each measure coded.

2.3.2 TOL scoring

Each participant was awarded a TOL Score. The efficiency of problem solving can be measured by the number of moves in which an item is solved, with no extra moves constituting the most efficient solving. The task comprised items of varying levels of difficulty, administered using a threshold procedure (see Method section, 2.2.4.2). The nature of the threshold procedure means that some participants completed more trials than others. The scoring system was devised to take into account both accuracy (i.e., the number of trials that were correctly completed) and efficiency (i.e., the number, if any, of additional moves that were made). Scores constituted a total number of points, with fewer points corresponding to better performance; in this way, problems solved in their minimum number of moves earned zero points (making the best possible score zero) and problems solved correctly but not in their minimum number of moves scored the number of additional moves made (with incomplete moves earning half a point; see Table 2.2). If 20 moves were made on an item without it being solved, the item was classed as incorrect (a 'too many moves' error). Therefore, these problems earned 20 points. Correspondingly, 20 penalty points were awarded for any incorrectly solved item, or any item in the task that had not been reached, because of the threshold procedure (this allowed for the number of correct items, as well as item efficiency, to be taken into account). This is because later problems were assumed to be more difficult than earlier problems, due to their increasing numbers of minimum moves as well as other difficulty-related parameters (see Appendix A for the full problem set). As there were four 2-move problems in the set and two of each of the 3,4,5 and 6-move problems, 2-move problem scores are half-weighted in the scoring. The score comprised the sum of the scores for each of the 16 trials. If more than half of the 2-move problems were incorrect, four 1-move problems were administered. The highest (i.e., poorest) possible score was therefore 240, assuming incorrect answers to all of the 2-move problems (and thus penalty points assigned for all later problems) followed by incorrect answers to all of the 1-move problems. In reality, the poorest score achieved was 200 (correct answers in one move to all 1-move problems). As no participant who completed 1-move problems made any additional moves, no portion of the scores come from these items. A scoring example is given in Table 2.3 below.

Performance	**TOL Score**
2-move problems: three correct in 2 moves, one incorrect	(0 + 0 + 0 + 20/2) points (half-weighted)
3-move problems: one correct in 3 moves, one incorrect	(0 + 20 points)
4-move problems: one correct in 5 moves, one in 6 moves	(1 + 2 points)
5-move problems: one correct in 10 moves, one incorrect	(5 + 20 points)
6-move problems: both incorrect	(20 + 20 points)
Total: 98 points	

Table 2.3: Example of TOL scoring

In order to illustrate the TOL score distribution, histograms of the score for each group are presented in Figure 2.10.

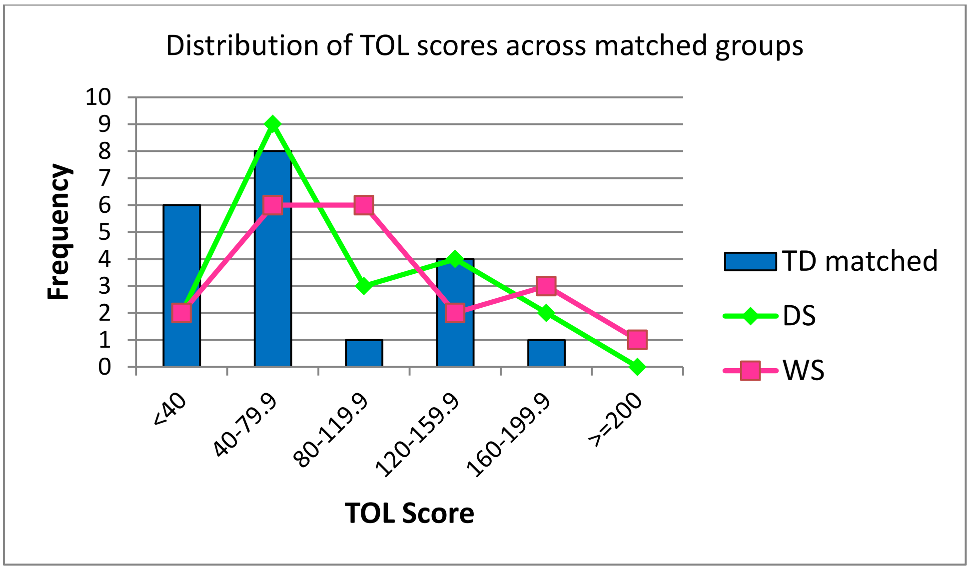

Figure 2.10: Frequency distribution of TOL score across matched groups

From examining Figure 2.10 the distribution of TOL scores appears to be broadly similar across the matched groups, with some unevenness between typical and atypical groups in the number of individuals scoring very well (less than 40) and those scoring towards the middle of the range (between 80 and 120).

Scores were also examined more closely by considering the two constituent measures: the trials correct score, which was the number of trials correct, half-weighted for the 2-move problems in the same way as for the TOL score; and the additional moves to correct trials, which was the average number of additional moves made on correctly completed trials, calculated for each participant.

2.3.3 Analyses and parametric assumptions

Participants completed a number of experimental tasks. Outcomes will initially be reported for the three RCPM-matched groups. Subsequently, the relationship between EF tasks and TOL score will be explored for each task for all participants. ANOVA analyses for group comparisons were examined using Tukey pairwise comparison tests when sample sizes were approximately equal and Games-Howell tests when they were unequal. The assumption of normality and homogeneity of variance for parametric analyses were checked prior to each analysis, using Kolomorov-Smirnov and Levene tests as appropriate. For group comparisons, where the normality assumption was not met, nonparametric tests were run, and reported if results were different from parametric equivalents. When the homogeneity of variance assumption was not met, the more robust Welch's F test is reported for between-group analyses. For trajectory analyses, where N was greater than 30 the assumptions of central limit theorem were adopted to allow an assumption of normality. Where Ns were below 30 and where necessary, the data were transformed to achieve normality. As reaction time (RT) data, having no maximum limit, are particularly vulnerable to the influence of outliers, analyses involving RT data were subject to the statistical identification of outliers, using a cut-off of two standard deviations from the mean. Such analyses are thus reported with and without the inclusion of outliers. Otherwise, accuracy data were included unless invalidated (e.g., by a lack of understanding of the task: details are given by task), as these are not as vulnerable to outliers as RT data and are likely to be valid given the wide variation in scores often seen in developmentally disordered groups. The one exception to this rule was for the proportional change in accuracy on the inhibition task where participants showing unusual patterns of behaviour were excluded: see footnote 2 on page 108. Where a score of zero was obtained on a task, the participant had demonstrated understanding of the task, without obtaining a score on experimental trials, and so the score was included.

2.3.4 Matched group comparisons

2.3.4.1 TOL Score

From the entire data set, data from five trials (1 TD, 1 DS, 3 WS) were excluded due to experimenter error, and missing values replaced with the accuracy score from the participant's remaining trial of the same number of moves. Four of those trials were from participants in the matched groups (1 DS, 3 WS). For the participant with WS, one of these excluded trials was a 2-move trial. As two out of the three remaining 2-move trials were correct, it was also judged as correct, and replaced with the average of the two correct trials' scores. On two additional trials participants from the two oldest year groups of the TD group were incorrectly given a more difficult trial. In both instances the trial was solved correctly. Scores are thus included, as any bias would be in a conservative direction. The mean scores for each group (N = 20 for each group) are displayed in Figure 2.11.

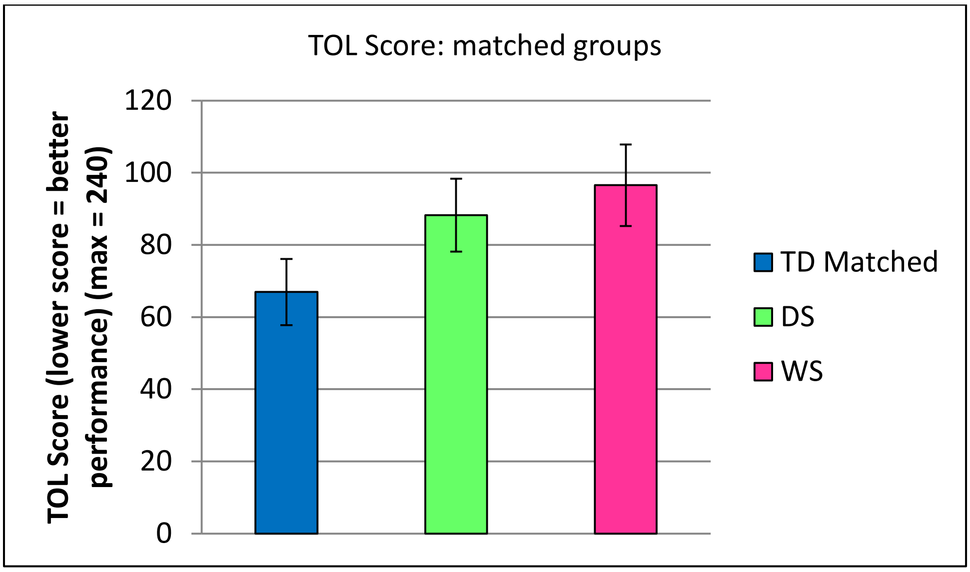

Figure 2.11: Mean *(S.E.)* of TOL score across matched groups. Lower scores indicate better performance.

ANOVA on the TOL score indicated no significant main effect of Group ($F(2,57) = 2.218$, $p = .118$, partial $\eta^2 = .072$). In order to compare both accuracy and efficiency, which are combined in the TOL score, across groups, ANOVAs were also conducted on the TOL trials correct score and the number of additional moves to correct trials. Neither measure indicated significant group differences (trials correct score: $F(2,57) = 1.928$, $p = .155$, partial $\eta^2 = .063$; additional moves: $F<1$). Means and standard deviations are displayed in Table 2.4.

	TD Matched	DS	WS
Trials correct score	9.40 (2.48)	8.40 (2.70)	7.73 (2.95)
Additional moves (correct trials)	1.48 (1.06)	1.87 (1.41)	1.53 (1.54)

Table 2.4: Mean (SD) of trials correct score and additional moves on correct trials

As the TOL task is the central task of the experimental work of this thesis, group comparisons were also conducted between each atypical group and the TD group for the performance variables above. This reduces the likelihood of not detecting differences due to a loss of power which may arise from comparing the three groups. Due to the centrality of the TOL task, this approach was only taken for this task.

The WS group performed more poorly than the TD group on the TOL score, ($t(38) = -2.029$, $p = .050$) which was largely due to the difference in the trials correct score which approached significance ($t(38) = 1.946$, $p = .059$) rather than the average additional moves to correct trials ($t(38) = -.133$, $p = .895$). None of the scores were significantly different between the DS and TD groups (TOL score: $t(38) = -1.557$, $p = .128$; trials correct score: $t(38) = 1.220$, $p = .230$; additional moves: $t(38) = -.993$, $p = .327$). While the reliable WS versus TD comparison would not have survived a Bonferroni correction we note it here because the correction is only warranted due to the multi-disorder group design. That is, had the study only included one atypical group and a control group, comparisons would have been made between those two groups only and would have been afforded more power.

A *post hoc* power analysis for the main effect of Group was computed using G*Power software (Faul, Erdfelder, Lang & Buchner, 2007), using an alpha level of

.05, effect size f of 0.28 (computed in G*Power from the partial η^2 = .072 above) and total sample size of 60 (three groups of 20). A power level of 0.45 was produced, that is, a 45% chance of detecting a violation of the null hypothesis. Cohen (1988) identified 80% as a desired level of power, so at 45% this analysis can indeed be considered as lacking power. Faul and colleagues (2007), amongst others, caution against the use of the sample effect size for *post hoc* power analyses, which cannot be assumed to represent the population effect size reliably. Thus, for completeness, a *post hoc* power analysis was also conducted based on detection of a medium effect size (f=0.25; Cohen,1988), which produced an observed power of 37%. From another perspective, an *a priori* power analysis for an ANOVA with three groups, an .05 alpha level, 80% power and a medium effect size (f=0.25) requires a sample size of 159, while the current sample was 60, so more participants would have been needed to detect a medium effect.

2.3.4.2 Planning task

The number of planning trials completed successfully was calculated for each participant. The practice trial required delivery to two houses, and some participants passed the practice trial but not the experimental trials, obtaining a score of zero. A further two participants (1 WS, 1 DS) were able to demonstrate a basic understanding of the task requirements by choosing the correct milk box for delivery to 1 house, even though they failed the 2-house practice trial. Their scores of zero were thus retained in the data set because they were able to understand the task at a basic level. One additional participant (with WS) was not given the opportunity to deliver to 1 house and so their data were excluded, so the Ns were 20 (TD), 20 (DS) and 19 (WS). Figure 2.12 displays the mean number of trials correct per group.

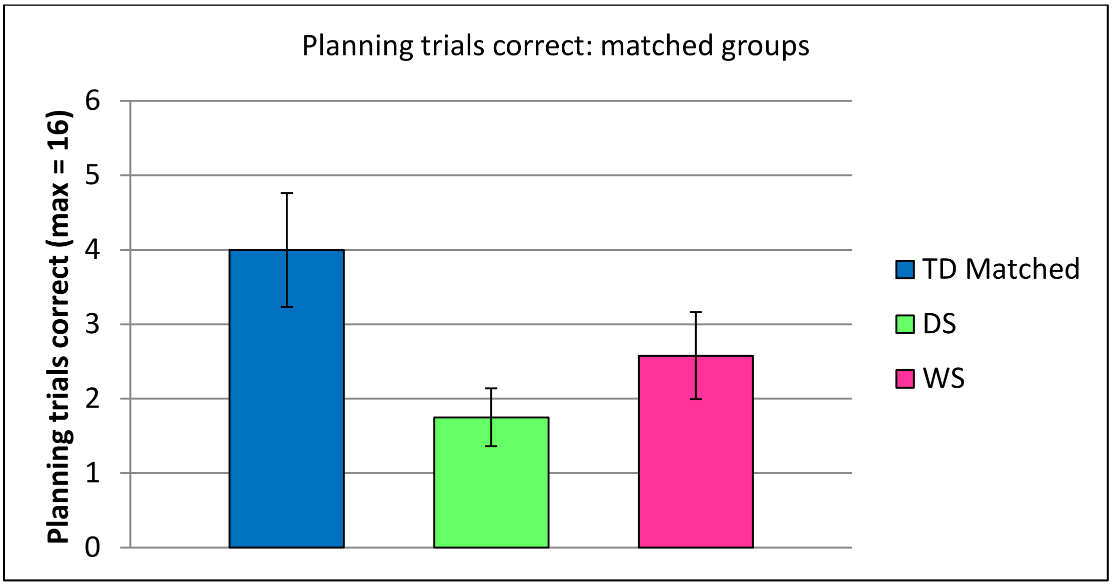

Figure 2.12: Trials correct on the planning task by group

ANOVA indicated a significant main effect of Group ($F(2,56) = 3.656$, $p = .032$, partial $\eta^2 = .115$), with better performance in the TD group than the DS group ($p = .026$) and no significant differences between the WS group and either of the other groups (Tukey tests: TD: $p = .227$, DS: $p = .597$). Planning data for the DS group did not meet the assumption of normality (Kolmogorov-Smirnov test: $p = .009$) which is likely to be due to the high frequency of a score of 1 in this group. A Kruskall-Wallis test of planning score across groups did not reach significance ($\chi^2(2) = 5.377$, $p = .068$), although *post hoc* Mann-Whitney tests produced the same pattern of results as the parametric tests above.

2.3.4.3 Shifting task

Data from three participants were excluded from the analysis (all from the matched groups) because they could either not understand or not perform the task (2 DS, 1 TD) leaving Ns of 18 (DS), 19 (TD) and 20 (WS). The group means of the number of shifts made per participant are displayed in Figure 2.13.

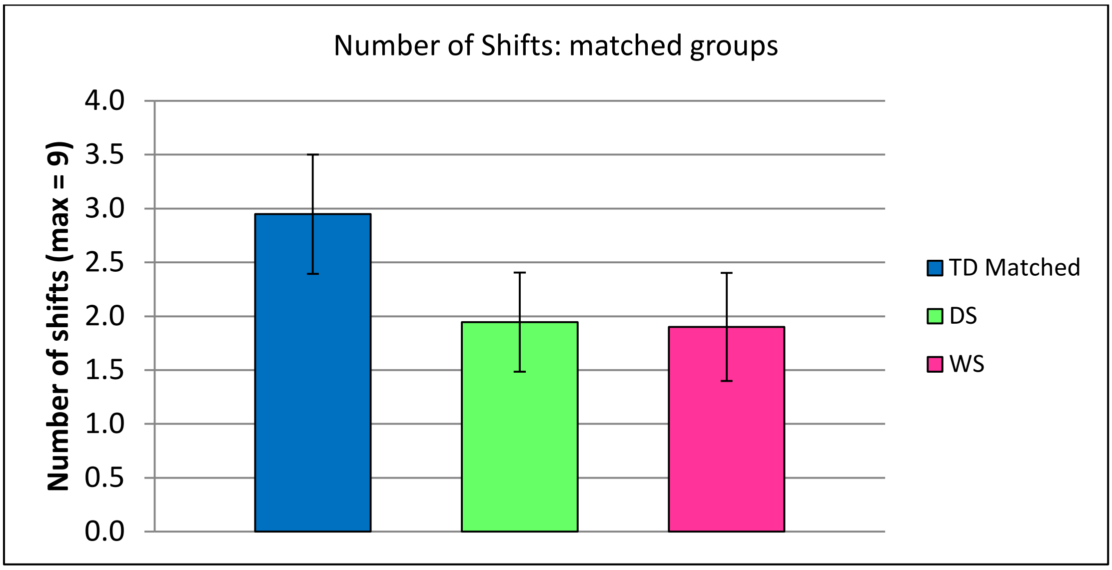

Figure 2.13: Number of shifts made on the shifting task by group. Stronger performance is reflected in more shifts.

There was no significant main effect of Group ($F(2,54) = 1.358$, $p = .266$, partial $\eta^2 = .048$).

2.3.4.4 Inhibition task

Accuracy and reaction time (RT) data were collected from this task.

2.3.4.4.1 Accuracy

The percentage of correct trials was calculated for each participant. Group means for each condition are displayed in Figure 2.14.

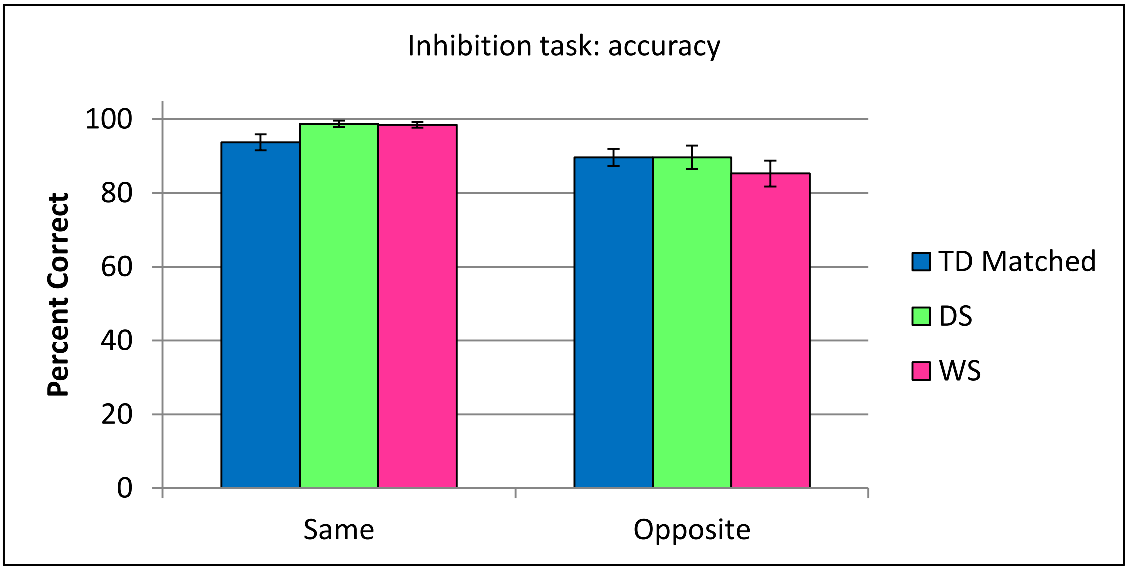

Figure 2.14: Percentage accuracy to same and opposite conditions of inhibition task by group

ANOVA with Group as the between-groups factor and Condition as the within-groups factor revealed a significant main effect of Condition ($F(1,57) = 26.537$, $p < .001$, partial $\eta^2 = .318$) with a higher rate of success to Same trials than to Opposite trials. There was no main effect of Group or significant Condition by Group interaction ($p > .05$ for both). Percentage accuracy data for the Same condition in the WS and DS group were very high, and did not meet the assumption of normality, while the homogeneity of variance assumption for this measure was also violated. One-sample t-tests against 100% performance indicated that these scores were indeed at ceiling (DS: $p = .163$; WS: $p = .056$). As accuracy was so high to the same condition it will not be considered further in relation to TOL score.

2.3.4.4.2 Reaction time (RT)

RT data points below 100ms were excluded, indicating either that a button press to a previous trial was still held down when a new trial began, or that a decision had been made by the participant regarding which button to press before the presentation of the stimulus. In the Same condition this affected three trials (2 TD, 1 DS). The average of the RTs across trials in the Same condition was calculated for each participant, and ANOVA on this baseline RT data revealed a marginally significant effect of Group ($F(2,57) = 2.712$, $p = .075$, partial $\eta^2 = .087$), with marginally longer RTs in the DS group than in the WS group ($p = .063$), with all other pairwise differences being non-significant ($p > .05$ for all). On excluding outliers above or below 2 standard deviations from the mean in each group (one participant in each group) the Group effect became significant ($F(2,54) = 3.650$, $p = .033$, partial $\eta^2 = .119$). Tukey tests indicated longer RTs in the DS group than the WS group ($p = .030$) with neither atypical group showing different RTs from the TD group (DS: $p = .726$; WS: $p = .160$).

In the Opposite condition, two trials were excluded due to RTs below 100ms (1 TD, 1 WS). The average of the RTs across trials was also calculated for each participant in the Opposite condition. The group means of RT for each condition are displayed in Figure 2.15.

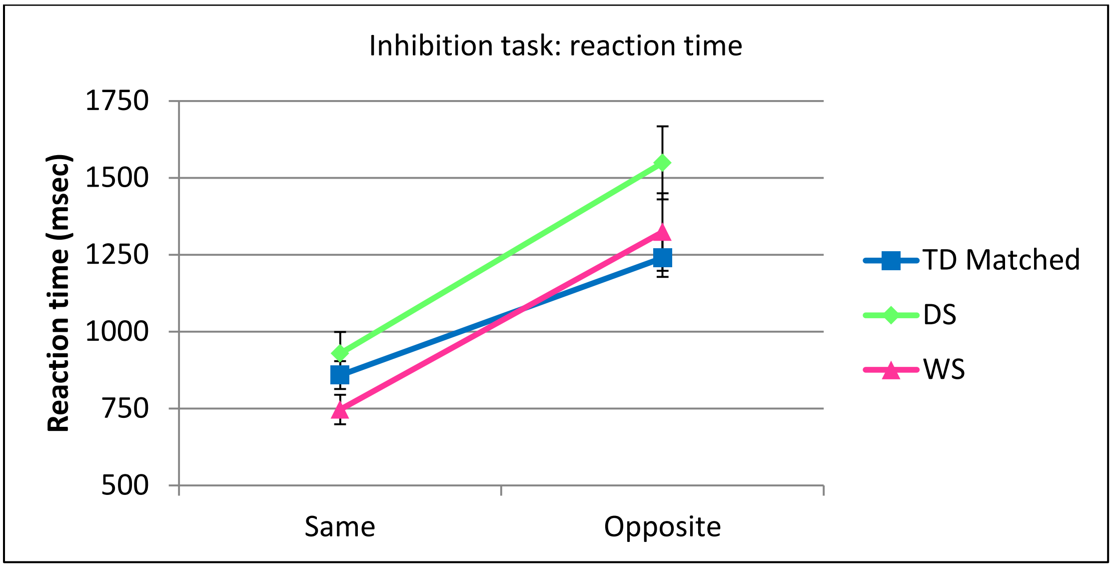

Figure 2.15: Reaction time to same and opposite conditions of inhibition task by group

ANOVA revealed a significant main effect of Condition ($F(1,57) = 100.986$, $p < .001$, partial $\eta^2 = .639$) with longer RTs to the Opposite condition than to the Same condition. There was a marginally significant main effect of Group ($F(2,57) = 2.507$, $p = .090$, partial $\eta^2 = .081$) with no significant differences between any of the groups in pairwise comparisons ($p > .05$ for all). Condition did not reliably interact with Group ($F(2,57) = 1.985$, $p = .147$, partial $\eta^2 = .065$).

In the Opposite condition, one participant in each group was identified as an outlier (these were different participants from those identified for the Same condition) and thus as being unrepresentative of the participant group, and were removed from the analysis. On excluding the outliers for both conditions, the main effect of Group became significant ($F(2,51) = 3.942$, $p = .026$, partial $\eta^2 = .134$), with Tukey tests indicating that the DS group's RTs were longer than the WS group's ($p = .021$) with neither atypical group performing differently to the TD group (DS: $p = .182$; WS: $p = .596$).

The proportional change in RT between conditions, relative to the baseline measure of matching speed (RT to the Same condition) was also calculated for each participant, using the following formula:

$$\frac{\text{Average RT Opposite} - \text{Average RT Same}}{\text{Average RT Same}} \times 100\%.$$

The group means (standard deviations) were as follows: TD 46.78 (25.80); DS 76.41 (56.87); WS 79.76 (59.06). There was a significant main effect of Group (Welch's $F(2,32.624) = 4.096$, $p = .026$, partial $\eta^2 = .086$) with a marginally significant difference between the TD and WS groups ($p = .099$) and all other differences being non-significant ($p > .05$).

Only one participant in the DS group was classed as an outlier for the RT Change variable. Upon excluding this outlier, the Group difference became significant (Welch's $F(2,32.341) = 3.535$, $p = .041$, partial $\eta^2 = .084$). Tukey *post hoc* tests indicated a marginally larger RT change in the WS than the TD group ($p = .078$) with no significant differences between the DS group and either of the other groups (TD: $p = .284$, WS: $p = .791$).

Thus, the results demonstrate that the WS group were marginally more impaired than the TD group when the inhibition of a match was required, with respect to RT but not accuracy.

2.3.4.5 Working memory tasks

There were four working memory tasks: forwards and backwards block span tasks and forwards and backwards digit span tasks. Data were available on all four of the tasks from 17 participants from the DS group, 18 from the WS group and 10 from the matched TD group (data from three participants with DS were unavailable for the backwards block task as they were unable to understand the demands of the task; remaining missing data are due to digit span measures only being tested for a subset of TD participants, and experimenter error). Mean span scores for the three groups on the four tasks are displayed in Figure 2.16.

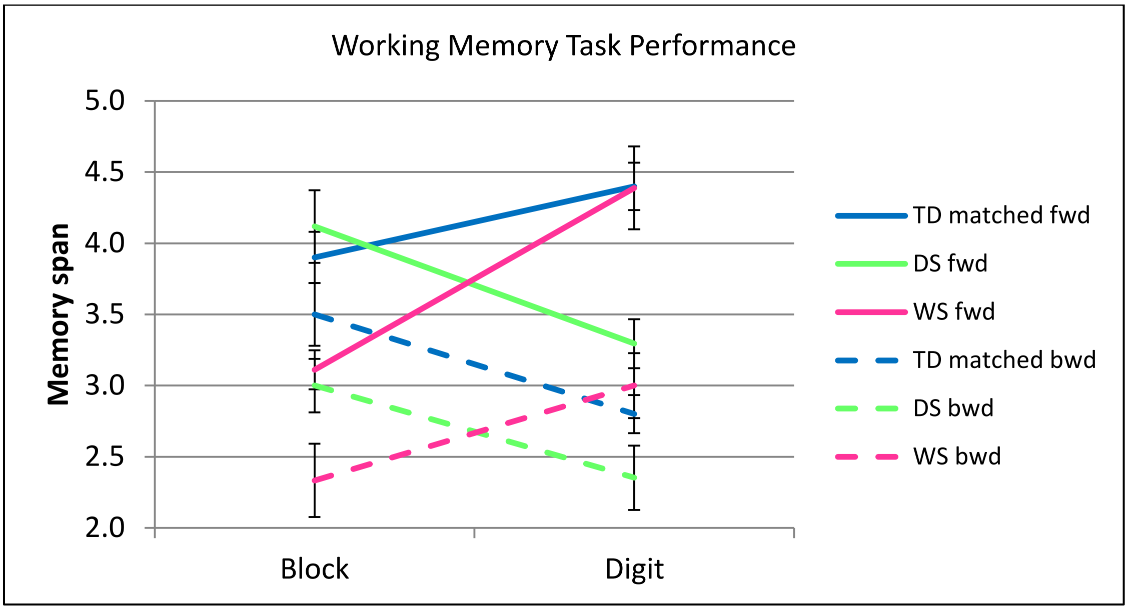

Figure 2.16: Span measures across four working memory tasks by group; fwd = forward; bwd = backward.

ANOVA was conducted with Group as the between-groups variable and Direction and Modality as the within-groups variables. There was no significant main effect of Group ($F(2,42) = 2.346$, $p = .108$, partial $\eta^2 = .100$) but its interactions with other independent variables are discussed below. There was a significant main effect of Direction ($F(1,42) = 83.037$, $p < .001$, partial $\eta^2 = .664$) with higher scores overall to forwards tasks than backwards tasks. Direction did not reliably interact with Group ($F<1$). There was no significant main effect of Modality ($F<1$) although Modality did interact reliably with Group ($F(1,42) = 20.414$, $p < .001$, partial $\eta^2 = .493$). The Direction by Modality interaction was also significant ($F(1,42) = 5.794$, $p = .021$, partial $\eta^2 = .121$) and these two variables contributed to a marginally significant 3-way interaction with Group ($F(2,42) = 2.966$, $p = .062$, partial $\eta^2 = .124$).

In examining the interaction between Modality and Group, one-way ANOVAs of block and digit spans, collapsed across direction, revealed main effects of Group for each (Block: $F(2,42) = 10.301$, $p < .001$, partial $\eta^2 = .329$; Digit: $F(2,42) = 6.554$, $p = .003$, partial $\eta^2 = .238$). Games-Howell *post hoc* tests for the block span measures revealed poorer performance in the WS than both the TD group ($p <$

.001) and the DS group (p = .003) with equivalent performance between the latter two groups (p = .786). Games-Howell *post hoc* tests for the digit span measures revealed poorer performance in the DS group than in the TD (p = .005) and WS group (p = .009) with WS and TD groups performing at a comparable level (p = .931). Paired sample t-tests indicated significant differences between digit and block spans in the DS (t = -3.922, p = .001) and WS groups (t = 4.765, p = .001), but not for the TD group (t = -.452, p = .662). The difference was in opposing directions for the two atypical groups, with the WS group obtaining better scores on digit span than block span tasks, and the DS group showing the opposite pattern, as to be expected.

As there was also a significant interaction between Modality and Direction, and the 3-way interaction approached significance, it merited exploration. It is explored in relation to group effects in order to be most informative. The main effect of Group for the block span was significant in both directions (Forwards: p = .002; Backwards p = .008) while for the digit span measure it was only significant for the forwards span (Forwards: p = .001; Backwards: p = .093, Kruskall-Wallis p = .102). For block span, the WS group scored more poorly than both other groups on the forwards task (TD: p = .007; DS: p = .005) but only poorer than the TD group on the backwards task (p = .005). For digit span, the DS group's performance was poorer than that of the other two groups on the forwards task (TD: p = .003, WS: p = .005). All other pairwise differences for the four tasks were non-significant (p > .05 for all); thus, for both backwards tasks the atypical groups' performance did not differ significantly.

The marginally significant 3-way interaction between direction, modality and group was also due to consistent effects of modality in each direction in the atypical groups, but inconsistent effects of modality in the TD group, when comparing forwards and backwards tasks. Paired samples t-tests revealed that the TD group performed better on the backwards block span task than the backwards digit span task (p = .010) with no such difference between the forwards tasks (p = .138). In the DS group, better performance to block tasks than digit tasks was seen in both directions (Forwards: p = .022; Backwards: p = .004) while the WS group did better at digit tasks than block tasks for both directions (Forwards: p < .001; Backwards: p = .029).

Overall, the expected pattern of block and digit span performance was observed in the WS and DS groups, with poorer visuospatial performance in the WS group and poorer verbal performance in the DS group, relative to both the other modality and to the other groups. These verbal/visuospatial differences between the atypical groups were only observed in the forwards direction, while within-group modality differences were observed for both directions in the atypical groups, and only in the backwards direction for the typical group.

2.3.5 Relationships between EF and TOL tasks

Thus far, comparisons have been made on various measures at the group level. This is informative with regards to the overall differences in performance across typical and atypical groups, for each task in turn. The next section considers the relationship between TOL performance and both MA measures and EF tasks, now for the entire TD group as well as each atypical group. First, correlations of EF task performance with TOL score were calculated for each group and developmental trajectories constructed and compared using ANCOVA (see Section 1.1.4) to assess the way, if any, in which the relationship between the EF task and TOL score differs across groups. Subsequently, multiple regression was used for the TD group to identify the most important predictors of TOL score.

2.3.5.1 Correlation matrix

The TOL score was correlated with each EF measure. This measure was selected because it is a more sophisticated measure of performance than the number of trials correct. The correlation matrix also includes the accuracy data equivalent of the proportional change in RT variable on the inhibition task, thus taking into account relative performance between conditions (labelled Inhibition: Accuracy Change in Table 2.5). This was calculated using the following formula:

$$\frac{\text{Average \% correct Opposite} - \text{Average \% correct Same}}{\text{Average \% correct Same}} \times 100\%.$$

In Table 2.5, the R values for group data that are non-normally distributed are underlined. Where this occurs and N is less than 30, Spearman's nonparametric correlations were also conducted. Where this outcome differed with regard to the presence of significance reported by the Pearson's correlation, the Spearman's R and *p* values are included in parentheses. Significant correlations are presented in Table 2.5 against a green background, while marginally significant correlations are against a yellow background. The Ns for each group (TD, DS, WS) are reported underneath the variable name in the leftmost column of the table.

	TD	DS	WS
	R (*p*)		
CA (56, 20, 20)	-.591 (<.001)	-.341 (.141)	-.187 (.429)
BPVS raw score (56, 20, 20)	-.554 (<.001)	-.287 (.221)	-.486 (.030)
RCPM raw score (56, 20, 20)	-.658 (<.001)	-.185 (.435)	-.278 (.235)
Planning score (56, 20, 19)	<u>-.473</u> (<.001)	<u>-.251</u> (.286)	-.682 (.001)
Shifts (54, 18, 20)	-.496 (<.001)	-.243 (.332)	-.331 (.154)
Inhibition: % Opposite (56, 20, 20)	<u>-.182</u> (.179)	-.322 (.166)	-.444 (.050)
Inhibition: Accuracy Change (56, 20, 20)	.185[2] (.180)	-.384 (.094)	-.469 (.037)
Inhibition: RT Same (56, 20, 20)	.457 (<.001)	.480 (.032)	.136 (.567)
Inhibition: RT Opposite[3] (56, 20, 20)	.532 (<.001)	.098 (.691)	-.411 (.080)
Inhibition: RT change (56, 20, 20)	.045 (.740)	-.386 (.092)	-.331 (.153)
Forwards digit span (26, 19, 19)	<u>-.531</u> (.005)	-.328 (.170)	-.352 (.140)
Backwards digit span (25, 19, 18)	<u>-.354</u> (.083) (-.496) (.012)	-.425 (.070)	-.447 (.063)
Forwards block span (54, 19, 20)	<u>-.353</u> (.009)	-.057 (.818)	<u>-.172</u> (.468)
Backwards block span (50, 17, 18)	<u>-.372</u> (.008)	-.002 (.993)	-.475 (.046)

Table 2.5: Correlations (r, with *p* value in parentheses) between predictor measures and TOL Score for each group. Significant correlations are presented against a green background whilst a yellow background indicates a marginally significant correlation. Underlined r values indicate non-normal data.

By examining Table 2.5 it is apparent that many variables were correlated with TOL performance for the TD group, as would be expected. While there was only one significant relationship with TOL score for the DS group (RT Same on the inhibition task), various aspects of the WS group's profile were associated with

[2] This correlation is after excluding two of the youngest participants who displayed unusual behaviour on the task: that is, better accuracy to the Opposite than Same condition. Before excluding them the correlation was significant (r = .373, *p* = .005).

[3] These correlations are after the exclusion of outliers (2 TD, 1 DS, 1 WS). Before exclusion, the WS correlation was not significant (r = -.146, *p* = .538) while the significance for the DS and TD groups remained unchanged (DS: r = .121, *p* = .612; TD: r = .572, *p* < .001).

TOL performance: spatial working memory and planning, vocabulary as measured by the BPVS, as well as inhibition.

As Table 2.5 contains 42 separate correlations, a Bonferroni correction for multiple comparisons would require a *p* value cut-off of .00119 for effects to be considered significant. If this were to be applied, correlations that would survive the correction would be: CA, BPVS, RCPM, Planning and Shifting for the TD group; none for the DS group; and only Planning for the WS group.

Correlations between percentage correct and RT in the Opposite condition were conducted to investigate the relationship between speed and accuracy in this task. Correlations were non-significant for each group ($p > .05$). However, on excluding the four outliers for the RT to the Opposite condition (1 DS, 1 WS, 2 TD; note, there is an additional TD outlier for the entire TD group than in the matched group analysis above) the TD and DS correlations remained non-significant while the WS correlation became significant (r = .590, $p = .008$), indicating an association between better performance on the opposite condition and longer response times for this group.

Correlations were also conducted between chronological age (CA) and the other variables in order to give an indication of which measures are changing over developmental time. The TD group showed significant correlations on every measure ($p < .05$ for all) with the exception of the proportional change in RT (r = .052, $p = .706$). Most of the correlations for the atypical groups were non-significant ($p > .05$), while the DS group showed age-related improvements in: BPVS; RCPM; Planning; and Accuracy on the opposite condition of the inhibition task ($p < .05$ for all) and in the WS group only BPVS score was related to CA (r = .546. $p = .013$).

Finally, some additional correlations between measures are reported here. Processing speed is an important factor in typical cognitive development (Fry &

Hale, 2000) and is indexed here by the RT to the Same condition of the inhibition task. In the TD group it was related to planning, shifting, all four memory tasks, BPVS, RCPM and CA, with faster RTs associated with better scores or increasing age ($p < .05$ for all). No correlations for the atypical groups reached significance, although one correlation approached significance in the DS group (backwards digit span: r = -.394, $p = .095$) and in the WS group (BPVS: r = -.425, $p = .062$).

2.3.5.2 Multiple regression: TD group

As the TD group has a large N, the sample size afforded multiple regression analysis to investigate the best way of predicting TOL score using the remaining variables. There is not enough evidence to confidently predetermine which variables will predict the TOL score most strongly: thus, a backwards stepwise regression method was chosen, to allow the selection of best predictors to be based on statistics rather than decisions based on theoretical considerations. The variables initially entered into the regression analysis were the ones that were correlated with the TOL score in the TD group: CA, BPVS, RCPM, RT Opposite, RT Same, Planning, Shifting, Forwards Block and Backwards Block spans, with the exception of the change in accuracy on the inhibition task (see footnote 2 on page 108). Digit span task sample sizes were not deemed large enough for inclusion at N=25 and N=26. However, for regression analysis multicollinearity should be avoided. This is defined (Field, 2013) as correlations between predictor variables with an r value greater than .9. The highest correlation was between CA and BPVS at .878. This is close enough to .9 to warrant concern. CA had the highest VIF value (variance inflation factor, an index of multicollinearity) of all the predictors (6.071) so was removed from the analysis. The resulting model with the highest R^2 value, thus accounting for the most variance in TOL scores, was the first model, including all the predictors, with an R^2 of .495, accounting for 49.5% of the variance. On removing each variable in turn, the value of F did not change significantly ($p > .05$ for all). RT Opposite and RCPM together still accounted for 46.9% of the variance. Thus, the RCPM and RT opposite variables explain a statistically equivalent amount of the variance as all the variables combined, and as such, this is the most parsimonious model ($F(2,45) = 19.867$, $p < .001$). The Durbin-Watson statistic was 1.565, which is within the normal range, indicating that the assumption of independence of errors was adequate (Field, 2013). Upon excluding the TD outliers for the RT Same and RT Opposite variables (N now being 52), the most parsimonious model now constitutes RCPM score in isolation ($F(1,43) = 26.922$, $p < .001$) accounting for 38.5% of the variance.

2.3.5.3 Developmental trajectories

ANCOVAs were conducted to assess the relationship between the EF or MA and TOL measure in each group, with TOL Score as the dependent variable, Group as the fixed factor and the predictor (MA or EF measure) as the covariate. An interaction term was included in the model (e.g., Group by Planning Score). To allow the comparison of the intercepts of the trajectories at the point at which they begin to overlap, measures were rescaled where necessary for analysis, by subtracting the highest minimum score across groups from each score. While a significant main effect of group indicates a significant difference in the intercept of the trajectories (that is, of the regression lines for the groups) a significant interaction points to unequal trajectory slopes for different groups, i.e., non-equivalent relationships between the measures for different groups. This section includes scatter plots of the raw data and the outcomes of ANCOVAs for variables in which at least two groups showed at least a marginally significant correlation between the TOL and the EF measure. A summary of outcomes for the remaining variables can be found in Appendix B. Ns for the DS and WS groups are identical to those in Section 2.3.4. The TD group N was 56 unless stated otherwise.

2.3.5.3.1 BPVS score

The BPVS raw scores and TOL scores are displayed in Figure 2.17.

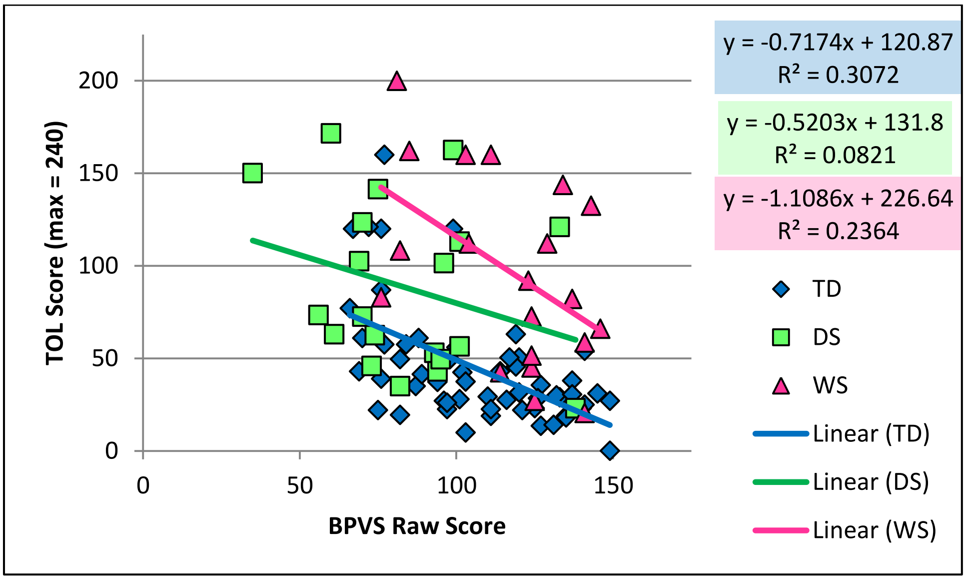

Figure 2.17: Developmental trajectory of TOL score based on BPVS score

There was a significant association overall between TOL score and BPVS ($F(1,90) = 20.105$, $p < .001$, partial $\eta^2 = .183$). Correlations revealed relationships between BPVS and TOL score in the TD and WS groups, but not the DS group. ANCOVA revealed that at the lowest level of verbal abilities, the TOL scores differed across groups ($F(2,90) = 9.126$, $p < .001$, partial $\eta^2 = .169$), with better TD scores than that of both atypical groups (DS: $p = .014$; WS: $p < .001$) as well as better DS than WS performance ($p = .045$) and the rate of TOL score development with BPVS score did not vary by group ($F<1$).

2.3.5.3.2 Planning task

Planning and TOL scores are displayed in Figure 2.18.

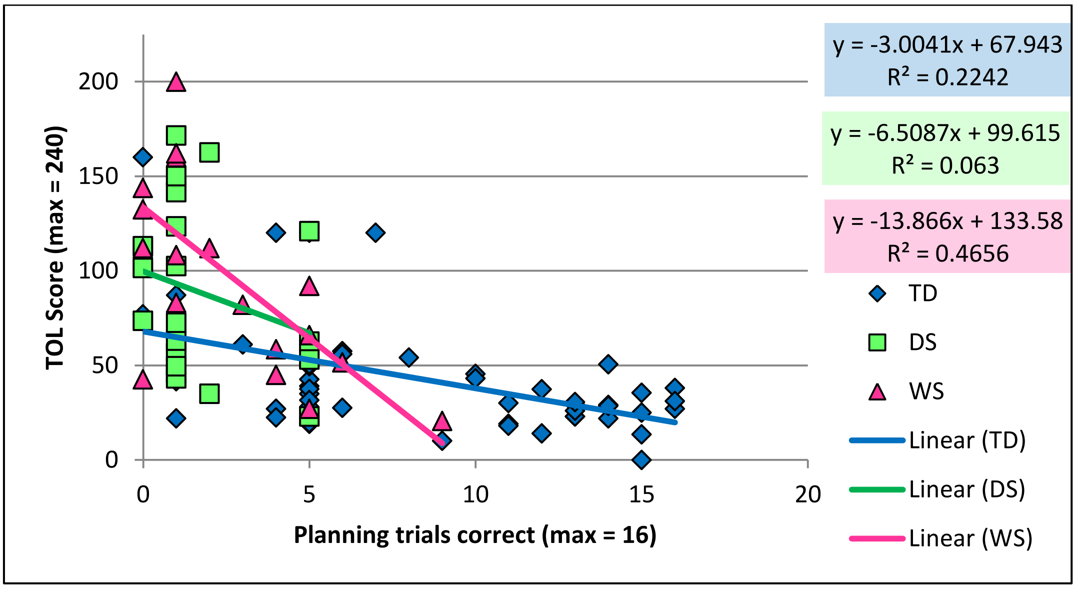

Figure 2.18: Developmental trajectory of TOL score based on planning task score

Planning data for the DS group did not meet the parametric assumption of normality (Kolmogorov-Smirnov test: p = .009) which can be attributed to the lack of variation of scores in this task in the DS group, with more than half scoring 1. Transformations did not improve normality to a satisfactory level. Thus, only data for the TD and WS groups were entered into the ANCOVA, which revealed an overall association between TOL score and planning score ($F(1,71) = 31.721$, $p < .001$, partial $\eta^2 = .309$) as well as significantly better TD than WS TOL scores at planning scores of zero ($F(1,71) = 25.753$, $p < .001$, partial $\eta^2 = .266$). Furthermore, the TD and WS slopes differed significantly ($p = .001$) indicating different rates of development of problem-solving ability with planning: in the WS group the slope is much steeper than that for the TD group, suggesting a greater improvement in TOL score with each additional planning trial correct; that is, that planning ability was a more important constraining factor for the WS group than the TD group.

2.3.5.3.3 Proportional accuracy change between conditions on the inhibition task

Figure 2.19 displays the scatter plot for the proportional change in accuracy between conditions on the inhibition task.

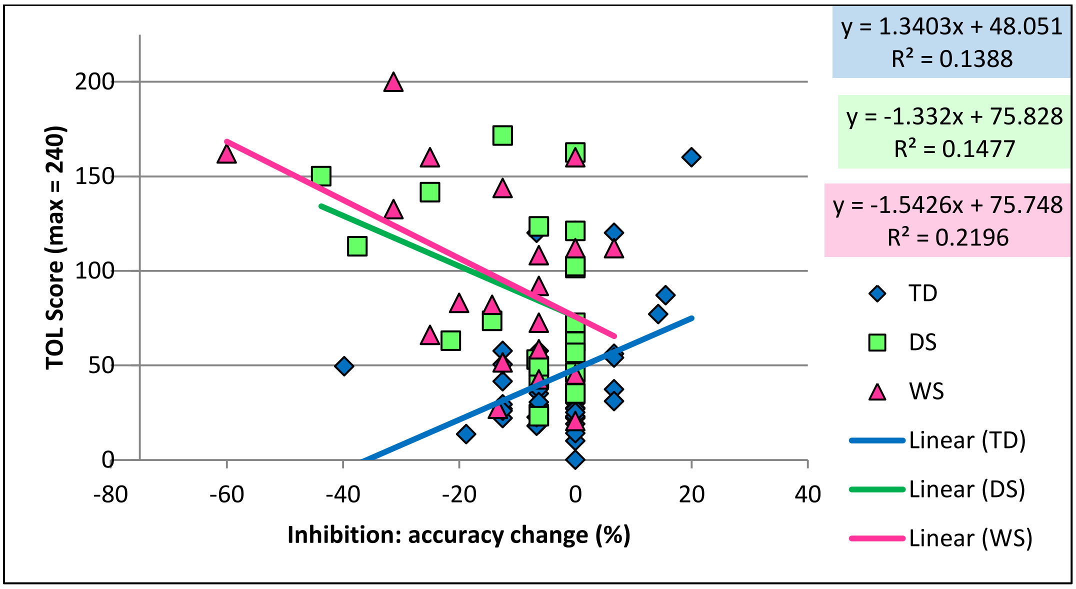

Figure 2.19: Developmental trajectory of TOL score based on proportional accuracy change on the inhibition task

As noted in Section 2.3.4.4.1, some of the TD children's accuracy was better for the Opposite than for the Same condition (those data points may be seen in Figure 2.19 where the scores are greater than zero). This measure was therefore not suitable for further analysis. Thus, only the DS and WS groups' trajectories were compared.

While the proportional change in accuracy was related to TOL score overall (p = .008), ANCOVA between the WS and DS groups did not reveal a significant difference in the change in accuracy at its lowest point, or a difference in the two groups' slopes (F<1 for both).

2.3.5.3.4 RT to the same condition of the inhibition task

In the inhibition task, reaction times (RT) to the Same condition showed significant relationships with TOL score in the TD and DS groups, but not in the

WS group, indicating that TOL performance was constrained by response speed in the TD and DS groups only. In Figure 2.20 the data delineating the relationship between these measures can be observed.

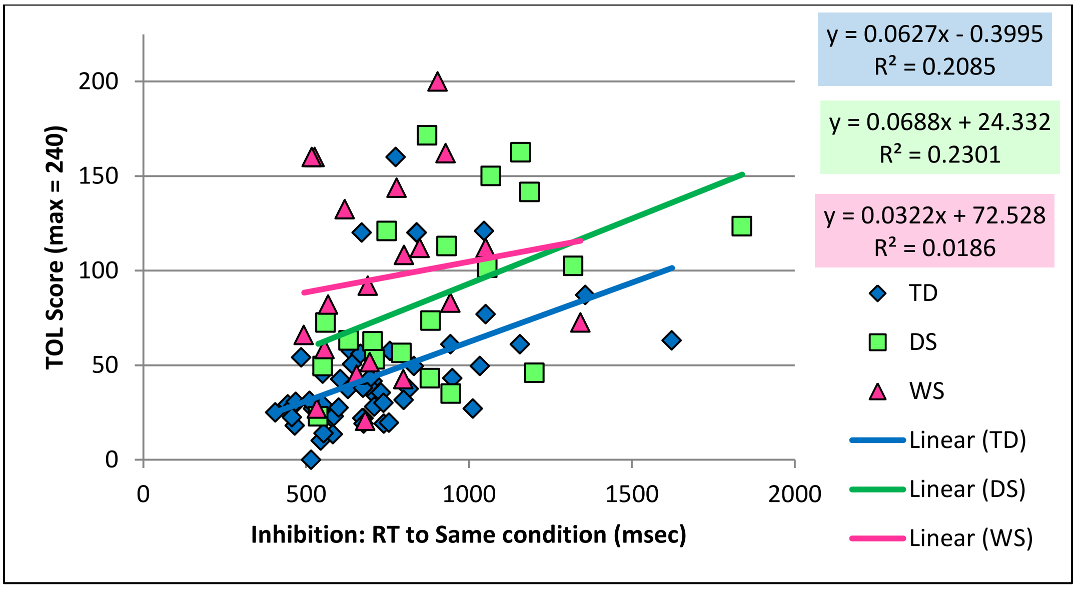

Figure 2.20: Developmental trajectory of TOL score based on RT to the same condition of the inhibition task

ANCOVA revealed that, at the fastest RTs, the three groups differed in their TOL score ($F(2,90) = 9.824$, $p < .001$, partial $\eta^2 = .179$) with TD scores better than those of the DS group ($p = .031$) and WS group ($p < .001$) and with no difference between DS and WS groups ($p = .210$). Overall, RT Same was associated with TOL scores ($F(1,90) = 9.856$, $p = .002$, partial $\eta^2 = .099$) and the rate of change of TOL score with RT did not differ across groups ($F<1$). Upon excluding outliers, the Group difference between the DS and TD groups lost significance ($p = .110$) with all other effects unchanged.

2.3.5.3.5 RT to the opposite condition of the inhibition task

While only the TD group showed a significant correlation between RT to the Opposite condition and TOL score, as noted in footnote 3 on page 108, upon excluding outliers, the WS correlation became marginally significant ($r = -.411$, $p = .080$). Interestingly this was in a different direction from the TD group.

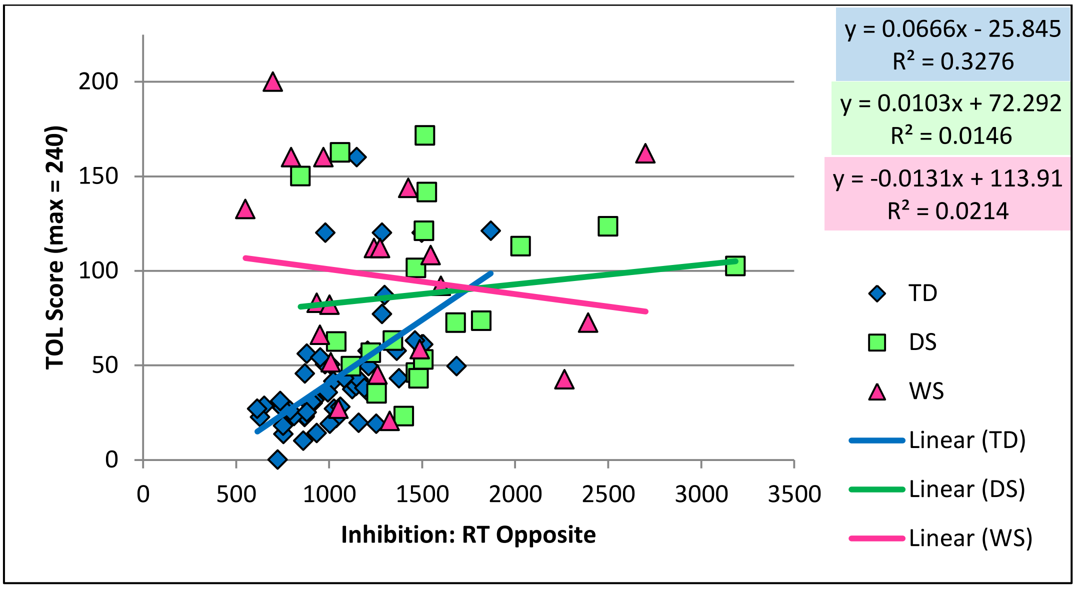

Figure 2.21: Developmental trajectory of TOL score based on RT to the opposite condition of the inhibition task

Figure 2.21 displays the scatter plot data for RT to the Opposite condition and TOL score. At the fastest RTs the groups scored significantly differently on the TOL ($F(2,90) = 18.942$, $p < .001$, partial $\eta^2 = .296$) with higher (i.e., poorer) scores in both the WS and DS groups compared to the TD group ($p < .001$ for both). While RT to the Opposite condition was associated with TOL score overall ($F(1,90) = 4.981$, $p = .028$, partial $\eta^2 = .052$), the relationship between the two scores was not consistent for all groups ($p = .004$). In fact, differences in slope were apparent between the TD group and both atypical groups (DS: $p = .010$; WS: $p = .001$). The two atypical groups did not differ in their scores at the fastest RTs or in their slopes ($p > .05$ for both). ANCOVA was also conducted with the outliers excluded. Overall, the association between RT Opposite and TOL score became non-significant ($p = .291$) while other effects remained unchanged.

2.3.5.3.6 Backwards block span

Data were available for 50 TD participants on this task. In Figure 2.22 a scatter plot of backwards block span and TOL score is shown.

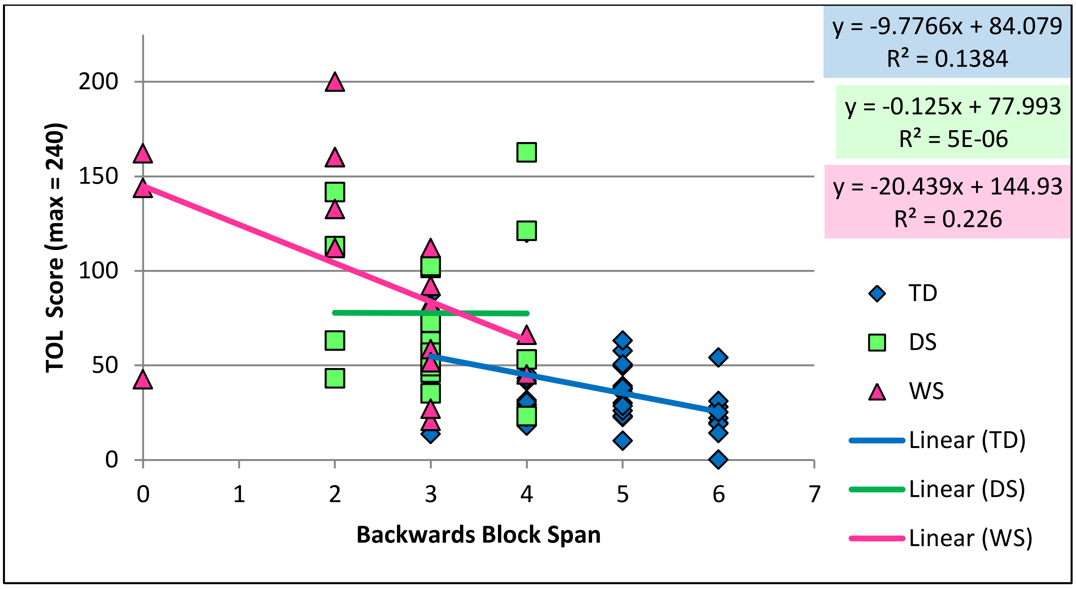

Figure 2.22: Scatter plot of backwards block span and TOL score

In the backwards block span task, significant relationships with TOL score were seen in the TD and WS groups only. This is unsurprising when examining Figure 2.22, as the DS scores only range from 2 to 4. In the remaining two groups, while both show relationships in the same direction such that better visuospatial working memory scores are linked to better TOL scores, the trajectories only overlap by one span point, and are thus unsuitable for comparison.

2.3.5.3.7 Backwards digit span

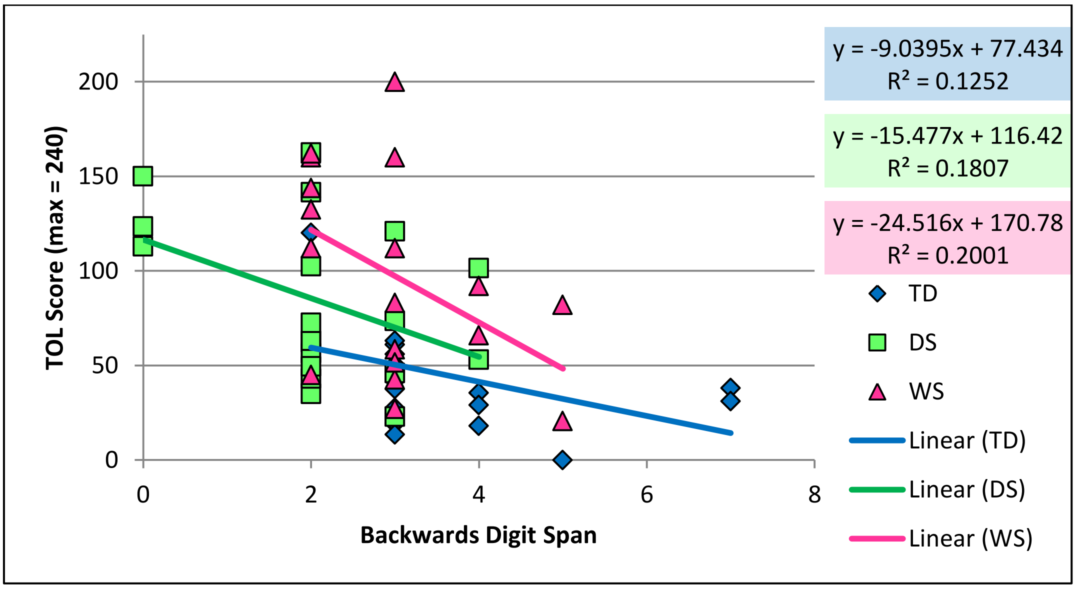

Figure 2.23: Scatter plot of backwards digit span and TOL score

In the TD group the distribution of backwards digit span was non-normal but the sample size was 25, negating the reliance on central limit theorem for an assumption of normality. Transformations did not improve normality and so the data were unsuitable for trajectory analysis. Figure 2.23 therefore serves to allow visualisation of the marginal correlation between the TOL and backwards digit span seen in all groups.

2.4 Discussion

The main aims of the study were 1) to compare EF performance in WS and DS to a control group matched on nonverbal ability, and 2) to investigate the relationship between EF (and MA) measures, and TOL problem solving, in WS and DS compared to TD individuals. TOL problem solving was found to be equivalent between RCPM-matched DS and TD groups but lower in the WS group than in the TD group. The patterns of EF performance of the three groups were also different. Moreover, different patterns of relationships between the TOL task and its predictor variables were also seen across groups, as expected, while some similarities were also observed. The nature of those differences and similarities, and relative EF profiles, will be discussed in this section. The reader is reminded that group comparisons are concerned with the RCPM-matched TD group while analysis of relationships between measures involves the entire TD group.

From a broad perspective, patterns of performance can be compared in three ways: relative group performance, the presence or otherwise of a relationship between a predictor variable and the TOL, and the nature of the development of that relationship, indicated by the developmental trajectory approach. The correlations between EF/MA measures and the TOL in the typical group serve as a benchmark for comparison for the atypical groups: if a measure is (or is not) related in typical development, and is not (or is) related in a disordered group, the presence or absence of that relationship may point to a discrepancy in the processes underlying TOL performance between those two groups. Having two atypical groups in the study also allows more information about that pattern of relationships to be known: is the unusual presence or absence of a relationship unique to a particular disorder or related to having a learning difficulty?

In the TD group, a predictor variable from every task was related to the TOL score, as was chronological age (CA) and both verbal and nonverbal MA. This is consistent with previous research yielding age-related increases in TOL score (Anderson et al., 1996; Culbertson & Zillmer, 1998; Huizinga et al., 2006; Luciana & Nelson, 1998) as well as being broadly consistent with studies finding some EF associations with TOL performance in TD children (Asato et al., 2006; Bull et al., 2004), although see, e.g., Bishop, Aamodt-Leeper, Creswell, McGurk, and Skuse (2001) regarding the Tower of Hanoi (TOH). These results vary by EF, and will be discussed individually subsequently.

For some predictor variables this relationship was only present in the TD group, while some were related in the TD group and in one atypical group, others were related in one atypical group alone and just one measure was related in all three groups. These will be discussed in turn, with reference to the nature of those relationships and to the matched group comparisons. The group comparison and developmental trajectory results will be further discussed for each task in turn.

2.4.1 TOL

The TOL score was significantly lower in the WS group than in the TD group matched for RCPM ability. The TOL score was a combination of accuracy and efficiency measures, and when comparing each of these constituent scores between the WS and TD groups it was suggestive that this group difference stems from lower TOL accuracy in the WS group than the TD group, although this difference did not reach significance. TOL performance in DS did not differ from that of the TD control group and can be understood to be at an equivalent level to performance that might be expected for TD children who exhibit the same level of nonverbal ability.

The slightly weaker TOL performance of the WS group compared to the TD group is consistent with previous studies by Menghini et al. (2010) and Rhodes et al. (2010) who found TOL deficits in a WS group compared to a TD control group. While scoring methods varied across studies, Menghini and colleagues also matched participants using a nonverbal measure of mental age (MA), the Leiter International Performance Scale-Revised, brief version (Leiter-R; Roid & Miller, 2002). Rhodes et al. (2010) used BPVS score as their group-matching measure. In the current study, the WS group's BPVS score was significantly higher, but their TOL performance significantly lower, than that of their RCPM-matched control group. Taken together, these findings serve to emphasise the elevated abilities on measures of concrete vocabulary seen in the WS population and the uneven cognitive profile (e.g., Mervis et al., 2000) previously found for this group. TOL performance appears to be not only lower than the WS group's relatively good verbal ability level, but lower than their nonverbal ability, when accuracy and efficiency are taken into account.

The current results for DS are consistent with those of Pennington et al. (2003) who found no difference in scores of DS and MA-matched controls on the CANTAB battery, including the Stockings of Cambridge (SOC) TOL-like task (also used by Rhodes et al., 2010 above). This is in contrast to those of Lanfranchi et al. (2010), in whose study adolescents with DS performed more poorly on a TOL test than controls matched on MA measured by a test of logical operations. However, their problem set consisted of 3- to 7-move problems, which is likely to have been more challenging than the current problem set of 1-2 to 6 moves. In addition, Rowe et al. (2006) found poorer TOL scores in a DS group compared to a BPVS-matched group of individuals with other learning disabilities, but as the score was based on the time needed to complete the items, it is not possible to meaningfully compare the scores (this study is further discussed in Chapter 3). A comparison of TOL performance in WS and DS was undertaken by Costanzo et al. (2013), comparing the atypical groups to a TD group matched on the Leiter-R. On a TOL task, a WS group obtained a lower score than a DS and TD group. Similar group patterns were also found in earlier studies assessing implicit memory in a DS or

WS group (Vicari et al., 2000, 2001). Here, the WS group also scored more poorly than the TD group, and the DS group scored at the same level as the TD group.

2.4.2 Chronological age (CA)

CA was related to the vast majority of measures for the TD group (without the proportional change in RT on the inhibition task), which is in line with expectations: EF skills increase with age in TD children (e.g., Best & Miller, 2010; Huizinga et al., 2006). While only BPVS score was related to CA in the WS group, both MA measures (BPVS, RCPM) as well as planning score and accuracy on the Opposite condition of the inhibition task were related to CA in the DS group. BPVS is a measure of crystallised intelligence, that is, concrete knowledge of words, so can reasonably be expected to increase with CA as individuals learn more words. This is consistent with Jarrold, Baddeley, Hewes, and Phillips (2001) who observed that, in WS, vocabulary developed faster than visuospatial construction skills (or at least it had an earlier onset age at which a score on the test is obtained; see Thomas et al., 2009 for a discussion).

Some data are also available in the existing literature regarding changes in IQ over developmental time in DS. IQ has been found to decrease during childhood in DS, but to remain largely stable across adulthood from age 21 to 45, when individuals affected by dementia are not taken into account (Carr, 2012). BPVS scores became more widespread over time, with those who were more able increasing in their scores the most (Carr, 2012), although gains were made between the ages of 21 and 30 for the group as a whole. It is therefore possible that, with age, increasing verbal skills for some of the participants with DS were related to increases in other abilities, although the age range in the present study is somewhat younger than that tested by Carr (2012), making comparisons difficult.

2.4.3 RCPM and TOL score

The Ravens test, of which the RCPM is one form, has been suggested to be ‘central’, because it shares a large amount of variance with performance on other tests (Carpenter, Just, & Shell, 1990). Therefore, it is thought to measure fluid, or analytic, intelligence, which does not depend on prior knowledge (Carpenter et al., 1990). The multiple regression analysis for the TD group revealed RCPM and RT to the Opposite condition of the inhibition task to be statistically just as able to account for TOL performance as a combination of all of the MA and EF measures used in the analysis, with RCPM alone coming through as the strongest predictor after the exclusion of outliers. Unterrainer et al. (2004) also found that fluid intelligence measured on a Raven's-like test was predictive of TOL performance in typical adults. Although in their study this was the sole predictor of performance and this was not the case in the present study, taken together, the Raven's tests were predictive of TOL ability in both children in the current study and adults in Unterrainer and colleagues' study.

The absence of a significant correlation between RCPM and TOL in the WS and DS groups serves to highlight the atypicality in the patterns of relationships between tasks and TOL performance for both of these groups; that is, there were other, more important constraining factors for performance.

Moreover, Carpenter et al. (1990) refer to completing the Raven's trials as solving problems, reflecting the wide applicability of problem-solving abilities, and the RCPM is also thought to involve EF skills (e.g., Rowe et al., 2006 included it as an EF task). Although it has also been suggested that the Raven's coloured matrices, suitable for children, can be solved perceptually (Hunt, 1974, cited in Carpenter et al., 1990), it is surprising that correlations between RCPM and TOL performance were not seen in either of the atypical groups. It might be questioned whether the RCPM test taps into abilities for individuals with neurodevelopmental disorders in the same way as for typical individuals. Adults with DS made different types of errors on the Raven's test compared to TD and learning disabled groups (Gunn & Jarrold, 2004). They, and another group with non-specific intellectual disability, also demonstrated different eye movement patterns to TD children during the task who were matched on mental age, while they did not differ from one another (Vakil & Lifshitz-Zehavi, 2012). However in Vakil and colleagues' study the TD group's Raven's score was better than the DS and intellectual disability groups', raising the possibility that different strategies (eye movements) facilitate better scores. Facon and Nuchadee (2010) found that items on the RCPM were comparable in their difficulty levels for a DS, intellectual disability and control group, indicating that this test is suitable for this type of comparison. In this study, groups were matched on RCPM score, indicating (unlike Vakil et al.) that the same scores appeared to be obtained for the same reasons across groups. Van Herwegen, Farran, and Annaz (2011) also matched a group of participants with WS to TD children on their RCPM score and a similar outcome to that of Facon and Nuchadee (2010) in DS emerged: error profiles in the WS group that were very similar to error profiles in typical children. These error profiles also changed with age in the same way for both groups, indicating the likelihood that the RCPM is measuring the same construct across groups (also see Facon, Magis, Nuchadee, & De Boeck, 2011 for further support for the use of the RCPM test to usefully compare typical and intellectually impaired groups). Here, RCPM performance was above the level expected for chance, indicating that responses were not being made at random.

It may be, then, that differences in the task demands between the TOL and RCPM rendered the association non-significant for the atypical groups. The TOL requires an individual to produce a motor response, and often, to make several steps towards the end goal, while the RCPM has limited motor requirements and only requires one response (although several internal steps may be taken before that response is produced). Indeed, the pattern of poorer performance on block design than RCPM performance in WS (e.g., Farran et al., 2001) indicates that, while both nonverbal, there are additional requirements to the block design task that make it more difficult. Similar processes may apply in the TOL.

2.4.4 BPVS

DS and TD matched groups did not differ on BPVS score, in line with Natsopoulo, Christou, Koutselini, Raftopoulos, and Karefillidou (2002) who found equivalent Raven's scores in DS and TD groups matched on verbal ability. In contrast, the WS group obtained higher scores on the BPVS than the other two RCPM-matched groups, which is consistent with the findings of Brock, Jarrold, Farran, Laws, and Riby (2007), and in line with the well-recognised discrepancy between verbal and nonverbal abilities in WS (e.g., Mervis et al., 2000; although this may not hold for all age groups; see Pezzini et al., 1999).

BPVS scores were related to TOL scores in the TD and WS groups, but not in the DS group. The BPVS measures verbal ability, using concrete vocabulary. The significant association in TD and WS groups might reflect the use of a verbal strategy, while the lack of significant correlation in the DS group suggests a dissociation between vocabulary and problem-solving abilities in this group. This is in line with Laws and Lawrence (2001) who did not find a link between verbal MA and drawing abilities in a DS group (while it was present in TD controls), and also with the suggestion by Laws (2002) of individuals with DS being less likely to use a verbal strategy than controls. Individuals with WS have also been suggested to use verbal mediation during tasks (anecdotally, Atkinson et al., 2003; Bellugi et al., 1994). However, given that the BPVS measures the knowledge of individual object-word associations, further research would be needed to elucidate whether a link between BPVS and TOL score necessarily points to the use of a verbal strategy (also see Section 3.3.4.3 for data regarding verbalisations during problem solving).

2.4.5 Planning

The TOL is largely known as a planning task (e.g., Shallice, 1982; Unterrainer & Owen, 2006) and has been found to be related to performance on the Porteus Maze Test, also thought to measure nonverbal planning (Krikorian et al., 1994) although the role of planning has been challenged in the TOL (Phillips, 1999; Phillips et al., 2001). Some versions of the task have required planning ahead, either in the task set-up or the instructions (Kaller, Rahm, Bolkenius, & Unterrainer, 2009; Klahr & Robinson, 1981) while others note that when not given instructions about strategies, people are likely to use a perceptual strategy which, rather than planning ahead, involves step-by-step 'online' planning (on the Tower of Hanoi, TOH; Goel & Grafman, 1995; see chapter 1 for more information regarding strategies).

In this study, the planning task was designed to require pre-planning, while the TOL procedure was deliberately set up to be a problem-solving task not necessitating pre-planning. As such, no information regarding the minimum number of moves required to solve a problem was given; no time limit was given; no instructions requiring people to plan, before executing their plan, were given. While self-correction was allowed in both the planning and TOL tasks (to a certain extent), the difference in the two tests lay in the requirement in the planning task to make all one's moves before any journey from start (van in centre) to goal (all boxes delivered), while in the TOL, moves could be made one move at a time, and thus plans could be made 'on-line' or updated throughout the execution of the task.

The WS and TD matched groups did not differ on planning scores, and correlations with TOL score were found for both groups, with better planning scores associated with improved TOL performance. At first glance, this pattern of results would suggest typical-looking planning processes in WS (in relation to MA). However the developmental trajectory approach affords analysis of the relationship between the two sets of scores between the groups, revealing significantly better TOL scores in the TD than the WS group from the outset of the developmental trajectory, as well as differences in the rates of development of problem solving with planning score. With each incremental increase in planning ability, the accompanying improvement in TOL score was greater for the WS group than the TD group, thus enabling them to essentially catch up to the TD group in their planning performance over developmental time. Indeed, while as a group, planning performance did not differ from that of the RCPM-matched control group, problem solving on the TOL was poorer in the WS than the TD group. These results together suggest that the planning score was a limiting factor for problem solving in WS at the lowest levels of planning ability, but that this does not wholly account for poorer TOL scores in relation to the TD group (see Section 2.4.6 below).

On the planning task, the DS group mean was significantly lower than that of the TD group. More than half of the DS group obtained a score of 1 on this task, indicating that they could successfully complete the experimental trial with two houses, but not three houses. The demonstration involved delivering to one house, and the practice trial to two houses. As many participants with DS could not deliver to three houses, this might indicate that many of them were unable to plan ahead by three steps, or could point to a difficulty with extending the strategy from the 1 and 2 house deliveries to trials with more houses. The lack of association between the two tasks in the DS group could be due to the low level of variance in planning score in this group. Alternatively, difficulties with planning might have meant that the DS group used an alternative step-by-step perceptual strategy instead of attempting to plan.

Poor DS performance on this task, taken alongside the equivalent TOL performance to the matched TD group, indicates that participants with DS were able to solve TOL problems to the same level as the TD group without the same capability to plan ahead. This is in contrast to the WS group, who were equally able to plan but less able to solve TOL problems than the controls. The different patterns of performance between planning and TOL scores found for both atypical groups (although in different ways) indicate that difficulties with planning ahead may be able to be successfully compensated for by individuals with DS, while individuals with WS are reliant on planning ability for problem solving at low planning levels.

In the current study, planning for the DS group, and TOL performance for the WS group, was the ability that showed the most impairment, being at lower levels than that of their MA-matched controls. The other of these two abilities (i.e., planning in the WS group; TOL performance in the DS group) was not significantly different to that of controls. It should be noted that equivalent performance to that of (often younger) MA-matched controls still indicates impairment relative to CA expectations; that is, both WS and DS groups showed impairment on both planning and TOL performance.

The differences seen in the patterns of performance across groups also speak to the planning/problem-solving distinction in the literature, suggesting that, at least for this particular TOL paradigm, the TOL was not simply a planning task (in line with intentions). This highlights the need to consider the particular administration of the task when identifying which abilities are important for the TOL. Had participants been told the minimum number of moves, or indeed only been allowed to make the correct number of moves, planning may have been drawn on much more strongly. Telling participants the minimum number of moves "focuses participants on the need for careful planning" (Berg & Byrd, 2002, p. 596). Planning times are longer when instructions to mentally plan first, and information about the number of moves, are given, than when they are not (Phillips et al., 2001). Whether this affects TOL performance is subject to debate (Phillips et al., 2001; Unterrainer et al., 2003).

2.4.6 Working memory

In the current study, forwards memory tasks were unsurprisingly completed more successfully than backwards memory tasks. As backwards tasks have additional requirements over forwards tasks (often recognised as measuring working memory, requiring manipulation of items, rather than simply short-term memory) this is consistent with expectations.

The interaction between Modality and Group reflects the pattern of findings existing in the literature. Poorer performance on the block span tasks in the WS group (both in comparison to the performance of the other two groups and in relation to their score on digit span tasks) and the same pattern of poor performance on the digit span tasks in the DS group, reflects the verbal/visuospatial unevenness seen in the cognitive profiles of the two disorders (e.g., Wang & Bellugi, 1994). Better spatial than verbal performance on the backwards tasks in the DS and TD groups suggests that spatial, rather than verbal, presentation appears to have been helpful. Visual prompting is likely to be a more helpful support strategy than auditory support in DS (Fidler & Nadel, 2007; also see Laws, 2002 regarding potential visual, rather than verbal, reliance for colour memory in DS) and visual prompts are generally known to reduce working memory demands in the typical population. For the DS group this spatial strength also applies (less strongly) to the forwards tasks. It might be that visual prompting only becomes helpful for the TD group on the more difficult backwards tasks, while the DS group struggle with verbal presentations regardless of task difficulty, so visual support reliably increases performance. Interestingly, the WS group showed the opposite pattern, with digit span tasks being easier than block span tasks in both directions. Thus, visual prompting may be a less helpful support strategy for individuals with WS.

WS performance was worse than that of typical controls for both forwards and backwards block span tasks, and DS performance was only worse than that of controls on the forwards version of the digit span task. Better WS than DS digit span and the reverse pattern for block span was seen for forwards but not backwards tasks, suggesting that the requirement to reverse the items in working memory rendered syndrome-specific differences undetectable.

The working memory findings in WS and DS are broadly consistent with existing literature, with some discrepancies that are likely to be due to differences in matching measures (e.g., Sampaio et al., 2008), tasks used (e.g., Carney et al., 2013) or in scoring methods (e.g., Costanzo et al., 2013; Menghini et al., 2010). For example, in Costanzo et al. (2013)'s study, WS and DS groups were also compared on forwards and backwards block and digit memory tasks. Both groups were poorer than controls on the block span tests, while only the WS group was poorer than controls in the current study; Costanzo and colleagues' DS group were poorer than the WS group for both forwards and backwards digit tasks, with worse WS than TD backwards performance. The between-syndrome findings are partially consistent with the current findings of poorer DS than WS performance to the forwards, but not backwards, test while no significant differences were found in the current study on the backwards digit span test. One reason for these discrepancies could be that Costanzo and colleagues used a scoring system that takes the number of correctly repeated sequences into account as well as the highest span achieved.

For the WS and TD groups only, better scores on the backwards block span task were related to better TOL scores. Whilst trajectories were unsuitable for comparison, the combination in the WS group of low ability on this task and correlation with TOL performance suggests that visuospatial manipulation of items in working memory may be a limiting factor for TOL problem solving. Potential reasons for the lack of correlation in the DS group could be a low level of variance in DS scores on the task (ranging from 2 to 4), or alternatively a difference in problem-solving style: if people with DS are not attempting to manipulate items in working memory to help them with the task (e.g., visualising where pieces will be placed) then a lack of correlation would be expected. However, given the correlation between backwards digit span and TOL score in this group, this explanation seems unlikely.

While the relationships between backwards digit span and TOL score should not be interpreted too strongly due to their marginally significant nature, they are interesting because this was the only measure to demonstrate a relationship with the TOL for all three groups. Significant performance differences on this task were also not seen between groups. This suggests that something in the requirement to reproduce a sequence of digits in reverse may have been inherent in the ability to solve TOL problems. The candidates for this could either be reversing items in memory or remembering sequences of digits. As forwards digit span, requiring the latter without the former, was only related to TOL score in the TD group, the most likely candidate seems to be reversing items in memory. One plausible interpretation of why reversing items in memory might contribute to TOL performance might be scenarios in which items need to be temporarily placed in a different order to that which is eventually required in the goal state.

The presence of a relationship between working memory (i.e., backwards span) measures and TOL performance in TD children partially supports the suspicions of Bull et al. (2004) and Kaller et al. (2008), who did not find relationships between forwards digit span measures (storage) and TOL performance, and suggested that TOL performance may depend more on working memory requiring manipulation of items (e.g., backwards tasks) in young children. However, in contrast to their findings, in the current study TOL was also related to forwards STM measures in the TD group. One reason for this discrepancy could be the wider age range of the TD group in the current study, compared to the 4-5 year olds tested by Bull et al. (2004) and Kaller et al. (2008). If STM storage is more important for older children than younger children, this could account for the lack of relationship for forwards memory tasks previously found (indeed, some evidence for the importance of working memory for TOL performance in children aged 7-8 and above was found by Asato et al. (2006) and Lehto et al. (2003) while Bull et al. (2004) and Kaller et al. (2008) did not find this association for younger children).

In the planning task participants were required to use the visual information available to them on the track to help them construct a plan for the sequence of items that would need to be delivered. It is therefore likely that the task was at least partly dependent on being able to visualise the route around the track, and manipulate items in working memory. The planning and backwards block span tasks seem to share a component of being able to mentally visualise and manipulate visuospatial items. Indeed, Kaller et al. (2004) note that, "as the planning process in the ToL relies on working memory involvement … confoundations with planning performance are likely" (p. 463). The reliance on these two abilities for TOL problem solving in the WS group thus seems intuitive, suggesting that the two tasks might be tapping into a similar underlying ability. The association between TOL performance and both of these tasks for this group suggests that the type of ability that these two tasks tap into is a limiting factor for the WS group in problem solving.

However, the WS group scored more poorly than controls on the backwards block span task, but not on the planning task. Individuals with WS are known to find tasks such as mental rotation difficult, while manual rotation is easier (Farran et al., 2001). Both backwards block span and planning tasks can be solved using mental manipulation, but the planning task also allows some use of a manual manipulation strategy. In the backwards block task, items must be immediately reproduced in reverse order, while in the planning task there is a phase in which the participant is loading items on the van, and it would be possible to use gestures in order to help, e.g., pointing to the house that would be needed next. It is possible that the WS group used manual manipulation as a compensatory strategy, with positive effects on planning task performance.

The question then arises as to whether the relationships between the two measures and the TOL were due to the requirement to manipulate, due to a requirement to plan that is also present in the backwards block span task or due to difficulties with visuospatial skills *per se*. As backwards digit span was also positively related to TOL performance in the WS group, it seems unlikely that TOL difficulties are purely related to visuospatial skills. However, because the planning and the problem-solving task were both visuospatial, it is not possible to fully tease apart planning abilities from manipulating items in working memory. Future studies which also include either a verbal measure of planning ability or a verbal problem-solving task, would be able to start to tease these abilities apart: if planning in one modality was a limiting factor for problem solving in a different modality, we would be able to conclude that planning into the future was a limiting factor for problem solving in WS.

2.4.7 Inhibition

2.4.7.1 Group comparisons

On the inhibition task, the accuracy data showed that performance was better in the Same condition than in the Opposite condition, as expected, reflecting the additional demands of the Opposite condition compared to the baseline Same condition. The lack of Group effect or Group by Condition interaction indicated that the three matched groups performed at an equivalent level, and responded in a similar way for the two conditions. Furthermore, both atypical groups were at ceiling for the Same condition (although for the WS group the difference from ceiling approached significance), indicating that they were able to complete this condition with ease. The difference in proportional accuracy change between conditions did not reach significance between groups.

The pattern of results was similar for the RT data: RTs were longer to the Opposite condition than to the Same condition, as expected, with no significant condition by group interaction or group effect, although after excluding outliers, the DS group RTs were longer than those of the WS group. Baseline RTs were also longer in the DS group than in the WS group, excluding outliers. The change in RT measure exhibited a trend towards a greater proportional increase to the Opposite than Same condition in the WS than TD group. For the WS group only, accuracy and RT were correlated in the opposite condition of the inhibition task, such that better accuracy was related to longer RTs. Put another way, responding quickly (impulsively) was related to poorer outcomes.

Although the WS group scored just as highly as the TD matched group on the inhibition task and did not take longer to respond *per se*, the relative increase in RT from the Same to the Opposite condition was marginally greater in the WS than TD group. This pattern and the avoidance of impulsive responding for better performance are partially consistent with previous findings of difficulties with inhibition in WS, reflected either by longer response times to inhibition tasks, e.g., to the Opposite world test measuring verbal inhibition (Menghini et al., 2010) and to the visuospatial inhibition task of Carney et al. (2013) (although Carney and colleagues did not find this effect for the verbal task) or by more errors than controls when response times are equivalent, e.g., on Go-NoGo and Stroop tasks measuring visuospatial inhibition (Menghini et al., 2010). This suggests the use of a more deliberate strategy in the Opposite condition by this group to enable success. It is possible that difficulties with inhibition, often found in WS, mean that to compensate for this they need to take longer to complete the task in order to score correctly. They may need to draw on a verbal strategy in order to do this, which would thus take more time. As mentioned above, the use of verbal mediation in WS has been suggested previously (e.g., Atkinson et al., 2003).

For the DS group, responses were slow compared to the WS group, but unlike the WS group, this was not related to their accuracy, and might indicate a generally slower responding style (cf. Gibson, 1991, cited in Cody & Kamphaus, 1999) rather than a particular strategy.

The current study used a nonverbal inhibition task. The finding of equivalent DS performance to that of controls is consistent with the findings of other studies using nonverbal inhibition measures (Carney et al., 2013; Pennington et al., 2003), while weaker inhibition in DS has often been found in studies using verbal tasks (e.g., Lanfranchi et al., 2010), which reflects the cognitive profile of better visuospatial than verbal skills in DS. Indeed, Costanzo et al. (2013) found this pattern of impaired verbal but not visuospatial inhibition in the same study. However, this pattern is not always manifest: for example Carney et al. (2013) did not find DS impairments on the verbal version of their inhibition task, while Rowe et al. (2006) did find DS impairments on a motor inhibition task, although differences in tasks or matching measures are likely to account for this (e.g., Rowe and colleagues compared their DS group to learning-disabled controls).

2.4.7.2 Relationships between inhibition and TOL Score

2.4.7.2.1 The TD group

It was noted by Huizinga et al. (2006) and Miyake et al. (2000) that inhibition is likely to be drawn on when solving tower-based tasks when instructions are not given regarding which strategy to use (although it should be noted that this is with respect to the TOH task). This is because when not given such instructions, people tend to use a perceptual strategy (Goel & Grafman, 1995), that is, moving towards the perceptual similarity one step at a time, for which inhibition is needed in order to avoid making moves that appear to be helpful but are not. As such instructions are not given in the current study this would suggest that inhibition would be needed, especially in light of previous findings of inhibition demands in TOL tasks for young children (Baughman & Cooper, 2007; Bull et al., 2004).

The TD participants were able to complete the inhibition task with ease. RTs were longer in the Opposite condition than in the Same condition, so there was clearly an impact of the requirement to inhibit matching and choose the alternative image. TOL scores were related to longer RTs but not to better accuracy on the Opposite condition, which is unsurprising as RT is the more sensitive measure. However the RT to the Same condition, or the baseline speed of response, was also related to TOL score for this group, while the measure designed to index the impact of the inhibition condition over the baseline condition, the proportional change in RT between the two conditions, was not related to TOL score. Assuming that the proportional change measure was successful in indexing this difference, this suggests that both the correlation between RT Opposite and TOL score and between RT Same and TOL score stem from the response speed rather than from the requirement to inhibit. Because the correlation coefficient was greater for the Opposite condition than for the Same condition, this also suggests that the Opposite condition RT contributes something above and beyond simple response speed. It is possible that an additional contribution of extra attentional resources or working memory (remembering to press the different picture rather than the same one) accounted for the difference between scores on the two conditions, (perhaps above the difference due to the requirement to inhibit): rather, TD children may have been able to 'set' a rule at the start of the Opposite condition and remember to follow it (cf. Bishop et al., 2001).

Bishop et al. (2001) measured inhibition in children using the Opposite World test from the test of everyday attention for children (TEA-Ch). In this test, children are first required to say '1' when shown number 1 and '2' when shown number 2. Subsequently, they are asked to produce the reverse response (e.g., '1' when shown '2'). Although a verbal test, this has clear similarities to the present inhibition task. Tower of Hanoi (TOH) performance was unrelated to Opposite world test performance in Bishop's study. Although the TOH and TOL are not equivalent tests (Bull et al., 2004), Bishop suggested that the type of inhibition that the Opposite world test taps into is too simple to be compared to the type of inhibition required in the TOH: rather than shifting between goals and sub-goals on the tower task, children just need to "maintain a response set that involves doing the opposite of what is customary" (p. 555) to perform the Opposite world test. Indeed, the presence of a correlation between shifting and TOL score for the TD group in the current study supports Bishop et al. (2001)'s assertions.

The finding of RT to the Opposite condition, but not the change in RT between conditions, being related to TOL score is consistent with that of Bull et al. (2004), who found that when controlling for baseline speed of naming items in their inhibition task, the correlation between inhibition latency and TOL score lost its significance for young TD children (although inhibition was still related as problems became more complex). A link between response time, or processing speed, and TOL score in the group of typical children might well be expected, in light of previous literature: typical development of processing speed sees rapid increases throughout childhood and is linked to increases in fluid intelligence, via increases in working memory (Fry & Hale, 2000) (developing in a nonlinear fashion with respect to CA over the lifespan). Asato et al. (2006) also found some evidence of a link between processing speed (time to initiate a visually guided saccade) and TOL performance, but only in 4-move problems. Taken together with the present outcomes, these findings suggest that response speed is an important consideration to take into account.

2.4.7.2.2 The atypical groups

On the inhibition task, for the WS group there was an association between better TOL scores and: better accuracy to the Opposite condition; smaller drops in accuracy from the same to the opposite condition, measured by the proportional change in accuracy; and, marginally, longer RTs to the opposite condition, after excluding outliers. We also know that, also after excluding outliers, better accuracy and longer RTs on the opposite condition were related to one another for this group. Thus, it may be that the individuals with WS who were taking more time to respond correctly to the inhibition task were also the individuals who obtained better scores on the TOL task. Inhibition may be related to problem solving in WS in that by avoiding impulsive responding, individuals are more able to consider the appropriate response. As the inhibition task did not appear to successfully tap into inhibition abilities in the TD group, it is difficult to say whether the WS association is typical or not.

Interestingly, the correlation between slower Opposite RTs and better TOL scores for the WS group is in the reverse direction to that of the TD group, for whom faster Opposite RTs were associated with better TOL scores. In addition, there was a lack of association between RT Same and TOL score in the WS group, while this association was present in the TD group. As we have seen in the literature discussed above, processing speed is important for TD cognitive abilities, but baseline speed of response does not appear to have been important for problem solving in the WS group. This potentially reflects atypical cognitive processing in the WS group, or could mean that any association between faster processing and better scores in this group has been diluted through the potential strategy, discussed above, of deliberately slower responding in order to avoid impulsive responding and obtain better scores.

For the DS group, the relationship between RT Same and TOL score was not only present and in the same direction as in the TD group, with the trajectory analysis revealing similar relationships between the measures for both groups, but it was the only measure to show a statistically significant association with TOL score for this group. This seems to indicate that when many skills need to be combined, e.g., in order to solve a problem, speed of information processing was the most important factor for success for the DS group. This is consistent with previous findings of slow information processing in DS (Gibson, 1991, cited in Cody & Kamphaus, 1999). Interestingly, Karmiloff-Smith (1998) suggested that DS may reflect a "failure to progressively specialise or modularize as a function of development" (p. 391). If this was the case, a relationship between general processing speed, rather than other abilities, and higher-level task performance could reflect this.

Studies of processing speed and working memory in typical children, (all cited in Fry & Hale, 2000), have indicated that increases in processing speed with age seem to account for much of the age-related increase in articulation rate (Hulme et al., 1984) and that processing speed serves to mediate the relationship between articulation rate and memory span (Kail, 1992). Moreover, both spatial and verbal processing speed predicted both spatial and verbal memory span to a similar degree, so this appears to be modality-general (Chuah & Maybery, 1999). Thus, it is possible that, in the TD group, the relationship between the RT measures and the TOL are interrelated to working memory span: for example, faster processing facilitating faster articulation rates (covert rehearsal) and allowing higher working memory spans, which facilitates TOL problem solving. In DS, the possibility that verbal memory deficits can be attributed to slow articulation rates has been refuted, for example by Jarrold, Baddeley, and Hewes (2000), who found evidence against a link between memory span and articulation rates in DS: verbal STM was impaired in DS compared to both TD children younger than 7 (before rehearsal begins to occur developmentally) and to individuals with similar articulation rates. However, in the present study, a measure of processing speed (RT Same) and memory span (backwards digit span) were both correlated with the TOL score in the DS group (backwards digit span, marginally so) and were also weakly related to one another. This raises the possibility that slow processing speed in DS (seen in the slower DS than WS RTs on the inhibition task) is related to difficulties with verbal memory, which both impact on TOL score. However this does not explain why forwards digit span was unrelated to TOL score in the DS group. It should also be noted that backwards digit span (requiring manipulation of items in working memory, as well as verbal memory span) was weakly related to TOL score in all three groups. A slow processing speed/poor verbal memory account of TOL performance for the DS group is therefore unlikely to be the whole story, and is unlikely to be unique to DS.

The inhibition task is an example of an instance in which identifying and removing outliers affected the pattern of results (here, removing participants identified as outliers for the RT data brought the group comparison to reach significance). In general, dealing with outliers presents its own challenge, in which a trade-off comes into play between discarding data points which are unrepresentative and may be masking true effects, and including data points which there is no reason to suppose are erroneous. Here, the approach taken was to present the analyses both before and after the exclusion of outliers in order to make both scenarios available to the reader.

2.4.8 Shifting

The number of shifts that were made did not differ across the three matched groups. This is consistent with the findings of Carney et al. (2013), who compared both WS and DS performance to that of TD children.

For the DS group, this is also broadly consistent with the findings of Pennington et al. (2003), who found equivalent performance on tasks purported to measure pre-frontal cortex (PFC) functioning (but not hippocampal functioning) in a DS group compared to a TD group matched on overall MA (recall that neither RCPM nor BPVS scores differed between the DS and TD groups here). Unfortunately, their study did not include a specific measure of shifting. In those studies that have included shifting tasks (Lanfranchi et al., 2010; Rowe et al., 2006), DS groups have performed more poorly than their comparison groups. However, it seems that differences in task administration and group matching measures could account for these discrepancies. This is, to the best of the author's knowledge, the first study to specifically compare shifting abilities in a DS group compared to a nonverbal ability-matched TD group.

Similarly, impairments in shifting abilities in WS compared to controls was found by Menghini et al. (2010) and Rhodes et al. (2010). Again, matching measures and tasks differed to the current study. In Menghini et al. (2010)'s study, although visuospatial shifting was poorer than that of controls, verbal shifting was not significantly different from that of controls, matched on a nonverbal MA measure. It is therefore possible that the participants with WS were able to use a verbal strategy to help them with the shifting task in the current study. Rhodes et al. (2010) found poorer WS shifting performance compared to a verbal-MA matched group, which might be expected as the measure (part of the CANTAB battery) was primarily nonverbal. As the study in question used a verbal age-matched group and the current study involved a nonverbal age-matched group, and the shifting tasks are both presented primarily non-verbally, it might be concluded that the WS group's nonverbal shifting ability should be expected to be in line with nonverbal mental age, as found here, but lower than would be expected for verbal MA, as found in Rhodes et al.'s study. Further research would be needed to detangle the effects of different comparison groups and tasks.

Shifting was related to TOL performance in the TD group, but in neither of the atypical groups. Shifting has been found to be related to TOL performance in typical children previously (e.g., Bull et al., 2004) so this is in line with expectations. Regarding the atypical groups, one possible reason for the lack of relationship might stem from the relative differences in success with the task across groups. While matched group shifting performance did not differ, only 20% of participants with WS and 33% of participants with DS were able to make 1 shift or more in phase 2, while the others either did not pass phase 1 or did not make any shifts in phase 2, compared to 68.5% of the entire TD group. The later phases require the discovery and adherence to rules, which is also required in the WCST (i.e., the adult version) while the DCCS, developed for children, involves telling children the rule they now need to follow each time it changes. In the TOL, there are no instructions regarding when to shift your attention between different aspects of the task (e.g., the overall goal, individual subgoals). Therefore, it may be that the type of shifting required on the TOL was better tapped into by the later phases of the task. Alternatively, individuals in the atypical groups may not be drawing on shifting.

2.4.9 Chapter summary

In summary, different patterns of EF performance and relationships between CA, MA and EF measures and the TOL task have been identified across WS, DS and TD groups, with the developmental trajectory approach revealing some differences as well as similarities in relationships between measures. On the TOL, poorer WS but equivalent DS performance was identified compared to a nonverbal MA matched TD control group. Many EF and MA measures were related to TOL performance for the typical group, as expected. While backwards digit span was marginally related to TOL score in all three groups, fewer significant relationships were observed for the atypical groups: indeed for the DS group, only RT to the Same condition of the inhibition task showed a reliable relationship with TOL score, highlighting the importance of processing speed, which was also thought to be important for the TD group. There were rather more significant relationships for the WS group. As well as a significant association between BPVS and TOL score, the inhibition task revealed a potential role of impulsive processing, suggested by the slightly elevated impact of the opposite condition on RT and the association between speed and accuracy, and between RT and TOL scores. For the WS group there were also significant relationships between backwards block span and planning task performance and the TOL, suggesting a limiting role of visuospatial working memory, and revealing an atypical relationship between planning task scores and TOL performance such that when an individual was poor at the planning task, this had a stronger adverse effect on TOL proficiency than for individuals who were relatively good at planning. Interestingly, Rhodes et al. (2010) did not find associations between EF tasks (one of which was a TOL-like task) in a WS group, while they did for typical children. Here the WS group did show some associations between the TOL and other EF tasks. This may be due to differences in tasks used and measurements taken, and warrants further research to account for the differences in patterns of association. For the DS group, planning task performance was poorer than that of the matched TD group, but unrelated to TOL performance, which may have arisen from a low level of variance in planning scores for this group, or a reduced reliance on planning because it is impaired, which may have led to compensation. RCPM score was the strongest predictor of TOL performance in the typical group but was unrelated to

TOL scores in both atypical groups, emphasising the atypicality of the patterns of relationships in WS and DS groups.

CHAPTER 3: EXPLORING PERFORMANCE ON THE TOWER OF LONDON TASK

3.1 Introduction

In this chapter, ways of measuring performance on the TOL, beyond the scores used in Chapter 2, are presented. Performance on a number of facets of TOL performance, and behaviours exhibited during TOL problem solving, is examined and discussed, with the aim of using these more sensitive measures to understand the processes operating during problem solving in Williams syndrome (WS), Down syndrome (DS) and typical development (TD). Each of these measures is reviewed in this introduction section.

3.1.1 Problem types

One reason for the development of the TOL was to provide a set of problems of different levels of difficulty (Shallice, 1982). Therefore, an overall score is likely to come from scores on a combination of different difficulty trials. Depending on the study's administration, it may be possible to obtain similar overall scores by scoring differently on different types of problems. For example, when measuring the number of trials correct, a score of say, four, could arise from correct performance on one of each of 2-, 3-, 4- and 5-move problems, or from two 2-move and two 3-move problems being scored correctly. As the former involves the correct solving of more difficult problems it merits more acknowledgement of success than the latter. Group performance has been shown to vary according to problem type: for example, a WS group was found to make more moves than controls on 3-, 4- and 5-move problems, but not on 2-move problems (Rhodes et al., 2010).

This matter is further complicated when we consider the concept of difficulty in the context of the TOL. As we have already seen, difficulty in this task can manifest in a variety of ways, represented by the properties that different problems can have (see Section 1.2.2.3). Several authors have emphasised the inadequacy of the minimum number of moves as the sole indicator of task difficulty (e.g., Kaller et al., 2004; Berg et al., 2010).

Different processes are also likely to be required to solve problems with different properties (e.g., Kaller et al., 2004). For example, solving TOL problems with more counterintuitive moves (that is, temporary moves away from the goal state; see Section 1.2.2.3.1) demanded greater inhibition abilities in children (Bull et al., 2004) and were associated with longer preplanning times in adults (Phillips, 1999). As discussed in Chapter 2, the current problem set included problems that were carefully controlled for, or systematically varied, with respect to several parameters of TOL problems (see Appendix A). Therefore, there is good reason to investigate both relative performance on problems of different types, and the relationship between scores on different problem types and EF measures.

3.1.2 Scoring

Various ways of scoring the TOL have also been employed in a variety of studies (see Section 1.2.2.3.4). Berg and Byrd (2002) recommended using several measures of performance simultaneously when administering the TOL task. Hughes (2002) noted the value, but scarcity in existing research, of the inclusion of observational measures of performance on the TOL task (e.g., rule violations) alongside the scoring measures. Some scoring measures beyond measures of accuracy are thus addressed in this section and used in the subsequent analysis.

3.1.2.1 Timing

Timing measures are widely used in TOL or TOH studies (e.g., Kaller et al., 2004; Newman & Pittman, 2007; Spitz, Minsky, & Bessellieu, 1985). Planning time ('first move time') represents the time elapsing before solving proper occurs, sometimes used to represent the time from the presentation of the goal to the end of the first move (Berg & Byrd, 2002) and sometimes, to the beginning of the first move (e.g., Huizinga et al., 2006). Planning time may change with age: Asato et al. (2006) report shorter planning times ('initial thinking time') in children than adolescents or adults for the most difficult problems, although Culbertson and Zillmer (1998) did not find changes in planning time with age from 7 to 12 years.

Asato et al. (2006) noted that spending longer on initial thinking time was related to solving problems in fewer moves, and the authors point out that this is particularly true for children, who spent less time on initial thinking than the other groups and made more moves when problems were more difficult.

Planning time has also been found to be affected by different problem parameters: for example, it was found to increase in problems with more ambiguous goal hierarchies (with flat-ending goals being the most ambiguous, and tower-ending the least ambiguous; see Section 1.2.2.3.1); and in the presence of non-optimal alternative solutions (Kaller et al., 2004), while for typical adults, longer planning times in the context of instructions to plan one's moves fully did not improve TOL solution efficiency (Phillips et al., 2001). Longer planning time could either reflect difficulties with putting a plan together, or could reflect careful planning ahead (Berg & Byrd, 2002 and references therein). Berg and Byrd (2002) also note that combining measures can help to separate out these possibilities – e.g., by examining associations between TOL score and planning time (cf. Ward & Allport, 1997).

Execution time (variously called, e.g., 'move time', 'subsequent thinking time') is the time elapsing from the end of planning time to the end of the solution. In Culbertson and Zillmer (1998)'s study, for children aged 7-12 years planning time remained constant, but younger children spent longer than older children on execution time. Of course, motor speed differences could account for some variations in the time taken to execute moves. Some authors (e.g., Asato et al., 2006; Rhodes et al., 2010) were able to control for motor speed using a yoked control condition as part of the CANTAB battery of tests. Asato et al. (2006) found that children spent longer on execution time than adults or adolescents.

Together, planning and execution time combined give the overall time to complete a solution. Anderson et al. (1996)'s raw score was produced by awarding points based on the total solution time (more points for faster solving), and subtracting the number of failed attempts. Increases in score were found between 7 and 9, and 11 and 12 years of age in typical children. Using total solution times will of course mask the relative contributions of planning and execution time, while combining measures to produce scores can also make interpretation difficult (cf. Berg & Byrd, 2002). Fortunately, Anderson and colleagues (1996) also analysed constituent measures, e.g., planning time and the number of failed attempts. However, this is not always the case: from the studies reporting timing measures from the TOL in WS and DS groups, two reported total time measures (Costanzo et al., 2013; Menghini et al., 2010) and another assigned scores for time-based categories (Rowe et al., 2006). Only one divided total time into its constituent planning and execution times (Rhodes et al., 2010), revealing longer planning time in a WS and chronological age-matched (CM) group than a verbal age-matched (VM) group (who had a lower mean chronological age than the other two groups). These outcomes also varied by problem type: the WS group took more planning time for 2- and 3-move problems than the other two groups did, but not for 4- or 5-move problems (while for these problems, the CM group time was longer than the VM group time). Thinking time (once pieces had begun being moved) was longer in the WS group than both control groups for problems of each number of moves.

Studies comparing a WS group to controls matched on nonverbal ability did not find group differences in total times to solution (Costanzo et al., 2013; Menghini et al., 2010) while in Costanzo et al. (2013)'s study, longer total solution times emerged in their DS group compared to their WS and TD groups. Rowe et al. (2006) used a scoring system which awards points based on the time taken to solve a problem (similar to that of Anderson et al., 1996 above), and found that DS scores were lower (i.e., taking more time) than those of verbal-matched learning disabled controls.

Thus, timing measures can further our understanding of the processes operating during problem solving, although they will not provide definitive indications of such processes. WS and DS groups have previously shown different patterns of response on tower-based tasks with respect to timing, compared to control groups. Splitting total time measures into separate planning and execution components is likely to provide more information than using total duration alone.

3.1.2.2 Rule violations (RVs)

The TOL requires that several rules are followed during the task, e.g., only one piece must be moved at a time. Measuring the number of times such rules are violated thus forms another way in which TOL performance can be assessed. Numbers of RVs were at low levels on the TOH task in children: e.g., 0.358 on average for 7-8 year olds and 0.177 for 13-15 year olds (Bishop et al., 2001). The mean number of rule violations increased from 0.7 for 35-year-old adults to 3.2 for 80-year-old adults, for the participants who were able to solve the 5-disc TOH (Ronnlund, Lovden, & Nilsson, 2001). While these are greater numbers of rule violations for adults than for children, which is unexpected, this could have been due to differences in the tasks administered: Bishop administered several trials of increasing difficulty with three or four discs, while Rönnlund used one 5-disc TOH task. Descriptive rule violation data were not reported for participants who were unable to solve the task. However, it was noted that age accounted for around 2% more of the variance in rule violations when non-solvers were included in the analysis than when they were not. Age, however, lost its significance as a predictor when the effect of other variables (gender, block design performance, word fluency, recall) was analysed.

Culbertson and Zillmer (1998) found that the number of RVs did not change with age for a TD control group, but decreased with age for a group of children with ADHD aged 7-12 years, while children with ADHD made more RVs than typical controls of the same age range (Culbertson & Zillmer, 1998), and there is further evidence of more RVs being made in other clinical populations compared to controls, e.g., in frontal lobe dementia (Carlin et al., 2000). However in Riccio, Wolfe, Romine, Davis, and Sullivan (2004)'s study, adults with ADHD did not differ from control and clinical disorder groups on TOL measures (including RVs) and RVs were associated with IQ, perceptual organisation, processing speed and matrix reasoning for the groups combined.

Children aged 7 years who were identified as 'hard to manage' at age 4 were less accurate than controls on a Go-NoGo task measuring inhibition, and made more RVs than controls on the SOC task (analogous to the TOL), while SOC scores *per se* did not differ between the groups (Brophy, Taylor, & Hughes, 2002). Baughman and Cooper (2007), citing data from Waldau (1999) in which 3-4 year old children made significantly more RVs than 5-6 year old children, produced a computational model in which increases in inhibition skills with age can account for differences seen in the RVs made between the age groups, and suggest that the inability to inhibit straightforward (yet incorrect) TOL strategies plays an important role in young children's TOL performance, including the violation of rules.

3.1.2.3 Errors

Another interesting way of understanding performance on the TOL is to consider the type of error that is made. Not much attention has been paid to this in existing literature. One exception is Welsh (1991), who examined errors occurring during TOH problem solving, which led solvers away from the optimum path for the solution, finding that children of all ages (3-12) exhibited the same types of errors on each problem. In the current study, participants could fail an item by solving it incorrectly, by making more than the maximum number of moves allowed, or by choosing to end a trial that they were unable to solve. Differences in error patterns are likely to provide clues about the problem-solving processes that occur across groups, and the approach taken by the individual.

3.1.2.4 Further measures of TOL performance

Some further types of behaviour that are addressed in the current study during the TOL task are outlined here: two different types of move and a measure of verbalisations made during the task. While most moves simply move a piece from one peg to another, Berg and Byrd (2002) identified 'incomplete moves', which involve picking up a ball and putting it back on the same peg. It may indicate "impulsivity or lack of planning" (p. 601) and they note that this type of move occurs more often in children than in adults. To the best of the author's knowledge, this type of move has not been measured in atypical populations.

The 'backup move' involves moving a piece from peg A to peg B, and returning it to peg A on the very next move. This type of move was observed by Klahr (1985) on the dog-cat-mouse puzzle (designed to be analogous to TOH puzzles in which the order of piece placement to reach the solution was ambiguous) in children aged 45-70 months. Klahr noted that while backup moves do not change the net state of the puzzle, and are thus not helpful for moving towards the solution, they are nevertheless useful when a participant has realised that the first move was incorrect and can correct their move. Low levels of backup moves were found overall in Klahr's study.

Private speech, which guides one's own behaviour, is thought to decline with increasing age, as it becomes more covert, and becomes inner speech, which is an internalised form (Vygotsky, 1934/1987, cited in Lidstone, Meins, & Fernyhough, 2010). Winsler and Naglieri (2003) examined verbal problem-solving strategies on the Planned Connections subtest of the Cognitive Assessment System (CAS; Naglieri & Das, 1997) in children aged 5 to 17 years, and found that private speech became increasingly covert with increasing age. Given the relatively good verbal skills in WS and suggested verbal mediation in this group (Atkinson et al., 2003), it was possible that more verbal strategies would be observed in the WS group than in other groups.

3.1.3 Aims

The aims of this chapter are to explore TOL performance in WS, DS and TD groups. First, performance on different types of problems will be compared, and EFs correlated with that performance, to examine if and how the EF demands change with problem type. We can reasonably expect a greater EF demand for more difficult problems as they are likely to require more resources to solve them. Second, the WS and DS groups will be compared to the RCPM-matched TD group on the behavioural measures that were collected during TOL video coding (see Table 2.2). This goes beyond the scope of most TOL studies. Behaviours can tell us about the ways in which people approach TOL problems, and may give clues about why people fail and how problem solving can be supported.

We might expect more verbalisations to be made in a WS group, if verbalisations are helpful for performance, particularly with reference to the possibility raised in Chapter 2 of a verbal strategy being used in the inhibition and TOL task for this group. We would expect longer planning times to be associated with better performance, if participants are actively engaging in planning. As the DS group were slower in the inhibition task, we would also expect their execution time/time per move to be longer than that of the other two groups (i.e., slower processing speed). Finally, if incomplete/backup moves do indicate on-line planning, we might expect a positive correlation between them and time taken per move.

3.2 Method

The method for data collection was identical to that included in Chapter 2. Details of the participants can be found in Table 2.1.

3.3 Results

In Chapter 2, five TOL trials from the entire dataset were excluded (0.47% of attempted trials), with the scores replaced with that of the remaining trial of the same length (4 moves, 5 moves…) or, in the case of a 2-move trial, with the average of the remaining correct 2-move trials. Four of these participants (1 DS, 3 WS) are part of the matched groups. For those trials, all of the measures (planning time, rule violations…) were excluded and replaced in the same way here. Some additional trials were excluded for analyses pertaining to individual measures, for which details are given in each section.

3.3.1 TOL scores by problem type

TOL scores were compared across problem types. The TOL score is the sum of the scores for each of the 16 problems. In order to compare TOL performance for different problem types across matched groups, the total TOL score was broken down into a constituent score for each of the five problem types, half-weighted for

2-move problems as before, for each participant, such that the sum of all of these equals the TOL score for that participant. The matched group means of these scores per problem type was compared in a repeated-measures ANOVA, with problem type as the within-subjects factor (2-, 3-, 4-, 5- and 6-moves) and Group as the between-subjects factor (WS, DS, matched TD) and are displayed in Figure 3.1 below.

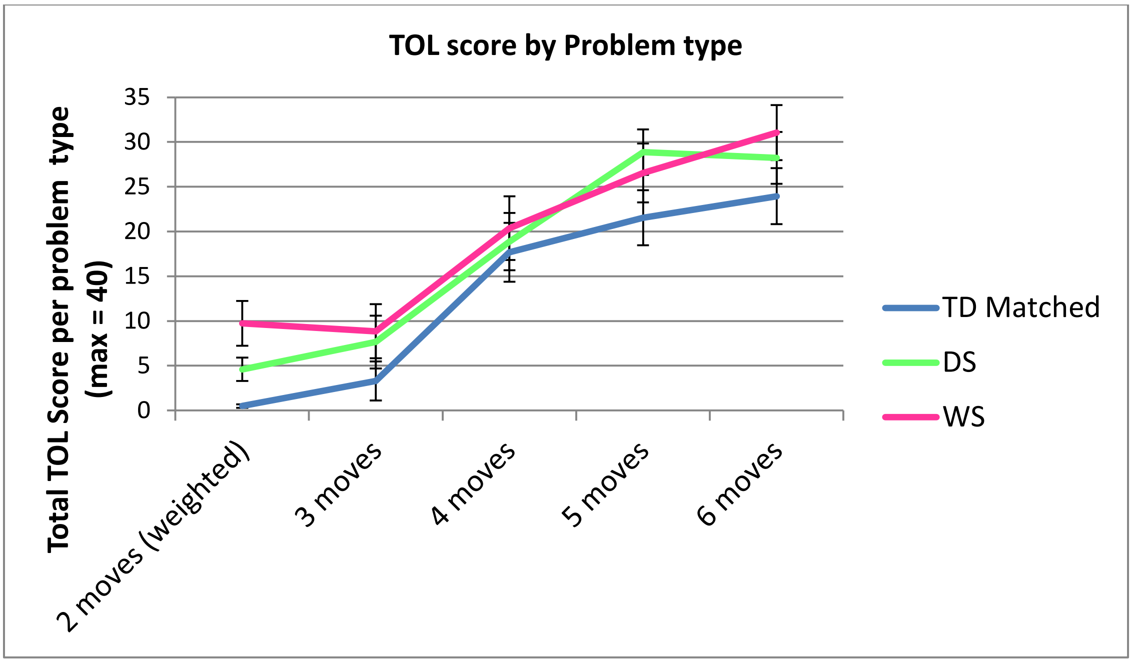

Figure 3.1: Group means *(S.E.)* total TOL score by problem type

ANOVA revealed a significant main effect of Problem type (F(4,228) = 72.215, p < .001, partial η^2 = .559). There was not a significant main effect of Group (F(2,57) = 2.218, p = .118, partial η^2 = .072) and Problem type and Group did not interact reliably (F<1). Pairwise comparisons revealed that 2- and 3-move problems (p = .294) and 5- and 6-move problems (p = .214) did not differ in their TOL scores, while 4-move problems were scored significantly differently from each other problem type (p < .001). Thus, 2-3 move problems were classified as 'easy', 4-move problems as 'medium' and 5-6 move problems as 'difficult'. Average scores were calculated for easy, medium and difficult problems for subsequent analyses. The trials correct score and average number of additional moves to correct problems reflected this pattern of performance, both showing better performance on easy than medium problems (p < .001 for both), and on medium than hard problems (trials correct: p < .001; additional moves: p = .017), with no main effect of Group (trials correct: $p = .243$; additional moves: $p = .115$) and no significant interaction between Problem type and Group (trials correct: $F<1$; additional moves: $p = .183$).

3.3.2 EF correlations by problem type

Having established three main types of problems, correlations of EF measures were then conducted for each group (the entire TD group), for each problem type, in order to ascertain whether different EFs are important for solving different types of problems. The correlation matrix is displayed in Table 3.1. Significant correlations ($p \leq .05$) are presented against a green background while marginally significant correlations ($.05 < p < .10$) are presented against a yellow background. R values for analyses in which at least one of the variables did not meet the assumption of normality (for the TD group, only when $N < 30$) are underlined, and Spearman's nonparametric equivalent correlations reported in parentheses if they produced a different pattern of results. The Ns for each correlation are included with the predictor name in Table 3.1 below.

	TD			DS			WS		
	Easy	Medium	Difficult	Easy	Medium	Difficult	Easy	Medium	Difficult
CA (months) (56, 20, 20)	-.231 (.087)	-.446 (<.001)	-.589 (<.001)	-.367 (.111)	-.288 (.218)	-.244 (.300)	-.245 (.298)	-.146 (.538)	-.090 (.707)
BPVS raw score (56, 20, 20)	-.168 (.217)	-.380 (.004)	-.594 (<.001)	-.292 (.211)	-.179 (.451)	-.262 (.265)	-.492 (.027)	-.404 (.078)	-.332 (.153)
RCPM raw score (56, 20, 20)	-.272 (.043)	-.505 (<.001)	-.646 (<.001)	-.336 (.148)	-.138 (.562)	-.034 (.888)	-.242 (.305)	-.437 (.054)	-.091 (.703)
Planning score (56, 20, 19)	-.210 (.120)	-.343 (.010)	-.472 (<.001)	-.204 (.389)	-.216 (.360)	-.231 (.328)	-.325 (.174)	-.712 (.001)	-.672 (.002)
Shifts (54, 20, 20)	-.255 (.063)	-.337 (.013)	-.504 (<.001)	-.025 (.922)	-.327 (.185)	-.258 (.301)	-.395 (.085)	-.330 (.156)	-.144 (.544)
Inhibition % Opposite (56, 20, 20)	-.130 (.341)	-.096 (.482)	-.188 (.166)	-.367 (.111)	-.269 (.251)	-.216 (.360)	-.459 (.042)	-.459 (.042)	-.238 (.313)
Inhibition: RT Same (56, 20, 20)	.195 (.150)	.395 (.003)	.397 (.001)	.430 (.059)	.417 (.067)	.405 (.077)	.147 (.535)	-.002 (.994)	.158 (.506)
Inhibition: RT Opposite (56, 20, 20)	.098 (.470)	.478 (<.001)	.586 (<.001)	-.049 (.836)	.163 (.493)	.191 (.420)	-.101 (.671)	-.191 (.419)	-.093 (.697)
Inhibition: RT change (56, 20, 20)	-.142 (.298)	.048 (.726)	.093 (.497)	-.436 (.055)	-.257 (.273)	-.308 (.187)	-.272 (.247)	-.264 (.261)	-.285 (.224)
Forwards digit span (26, 19, 19)	.030 (.886)	-.315 (.117)	-.592 (.001)	-.204 (.403)	-.288 (.232)	-.329 (.169)	-.426 (.069)	-.347 (.146)	-.151 (.536)
Backwards digit span (25, 19, 18)	-.105 (.617)	-.171 (.413)	-.410 (.042)	-.513 (.025)	-.427 (.068)	-.214 (.379)	-.243 (.331)	-.590 (.010)	-.339 (.168)
Forwards block span (54, 19, 20)	-.186 (.178)	-.310 (.022)	-.309 (.023)	-.033 (.893)	-.042 (.866)	-.068 (.784)	-.060 (.801)	.055 (.817)	-.334 (.150)
Backwards block span (50, 17, 18)	.004 (.980)	-.170 (.239)	-.439 (.001)	-.103 (.694)	.157 (.546)	-.034 (.898)	-.308 (.214)	-.503 (.033)	-.399 (.101) (.509, .031)

Table 3.1: Correlations (r, with *p* value in parentheses) of EF measures with TOL score for easy, medium and hard problems. Significant correlations are against a green background and marginally significant ones against a yellow background. Where data for at least one variable was non-normal, the r values are underlined.

Table 3.1 contains 117 separate correlations. A Bonferroni correction for multiple comparisons would require a *p* value cut-off of 0.000427 for effects to be considered significant. It should be noted that SPSS does not report *p* values below .001, merely stating that they have a lower value than .001. Taking this as the cut-off criterion, only the nine correlations for TD group with this value (Medium: CA, RT Opposite and RCPM; Difficult: CA, BPVS, RCPM, Planning, Shifting and RT Opposite) would survive the Bonferroni correction.

3.3.3 Timing measures

Timing measures were also compared across the three problem types. Planning time (from goal presentation to the first move) and total trial time (from goal presentation until the last move) were recorded – see Table 2.2. Measurements were taken from the video recordings using a stopwatch, rounded to the nearest second (for planning time) or from the software's timer when playing back the videos (for total trial time), which was accurate to the nearest second. The execution time was then calculated for each participant by subtracting the planning time from the total trial time. Execution time was divided by the number of moves made to obtain the time per move. Timing measures were calculated for each attempted trial for each participant, and averaged across trials of each problem type (easy, medium, difficult), half-weighted for 2-move problems.

Planning time and time per move were excluded for five trials (four from one participant with DS and one from one TD participant; both were included in the matched groups) due to equipment malfunction or inability to see from the video recording when the first move began. On three further trials, timing measures were excluded for one trial from a TD participant and two trials from a participant with DS who did not make a response to the trial (and thus planning time and time per move were not possible to measure). Timing measures were also excluded for those participants (2, TD) for whom the trial was incorrectly set up, one of whom was part of the matched group. Both timing measures were thus excluded from nine trials from the matched group data.

3.3.3.1 Planning time

Planning time was additionally excluded for one trial (TD) because the participant was distracted during the planning time window. This participant was not in the matched group. Planning times recorded as under 1 second were classed as zero, as they were most likely to reflect the time required to press the stopwatch rather than any pre-planning exhibited by the participant. This occurred for 23 trials, of which 8 were from participants included in the matched groups. This comprised 2.16% of all attempted trials, and 1.27% of all attempted trials for the matched groups. Within the matched groups, the Ns for comparison of the three problem types (i.e., those participants who attempted easy, medium and difficult problems) were 15 (TD), 14 (DS) and 14 (WS). Group planning time data are displayed in Figure 3.2.

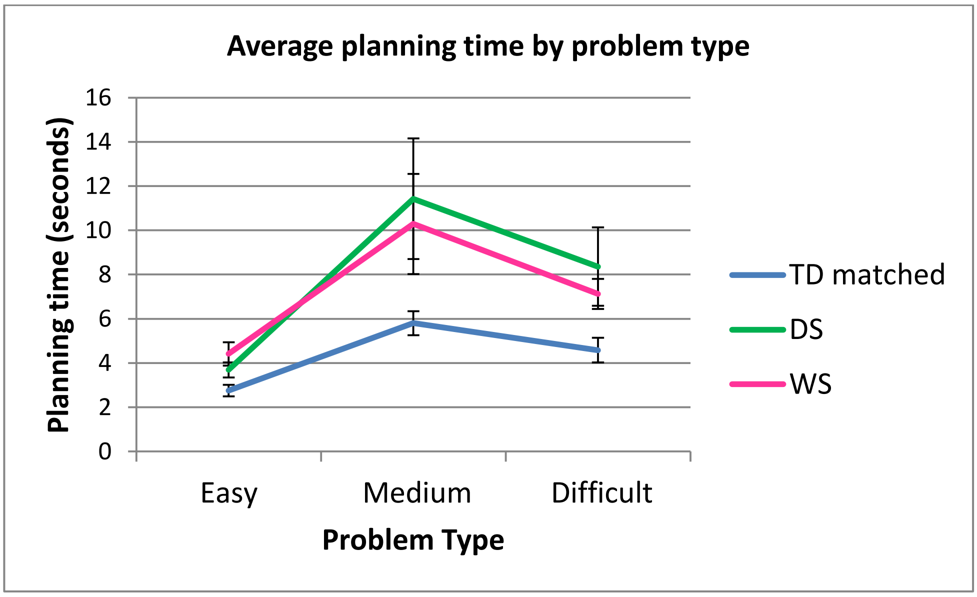

Figure 3.2: Group mean *(S.E.)* of planning time (seconds) by problem type

A repeated-measures ANOVA on planning time with Group as the between-subjects variable (WS, DS, matched TD) and Problem type as the within-subjects variable (easy, medium, difficult) revealed a significant main effect of Group ($F(2,40) = 3.296$, $p = .047$, partial $\eta^2 = .141$) with Tukey tests detecting marginally longer planning times in the DS than the TD group ($p = .057$) with no significant differences between TD and WS groups ($p = .126$) or DS and WS groups ($p = .926$). There was also a significant main effect of Problem type ($F(1.361,54.443) = 19.925$, $p < .001$, partial $\eta^2 = .332$), with significantly shorter planning times for easy problems than for both other types ($p < .001$ for both) and also significantly shorter planning times for difficult than medium problems ($p = .005$). The interaction between problem type and group was not significant ($p = .286$).

3.3.3.2 Time per move

Execution time data (and therefore time per move) were additionally excluded (beyond the nine trial exclusions outlined at the beginning of Section 3.3.3) for one participant with WS and one participant with DS because of a substantial interruption noted during testing. Data for group means are displayed in Figure 3.3.

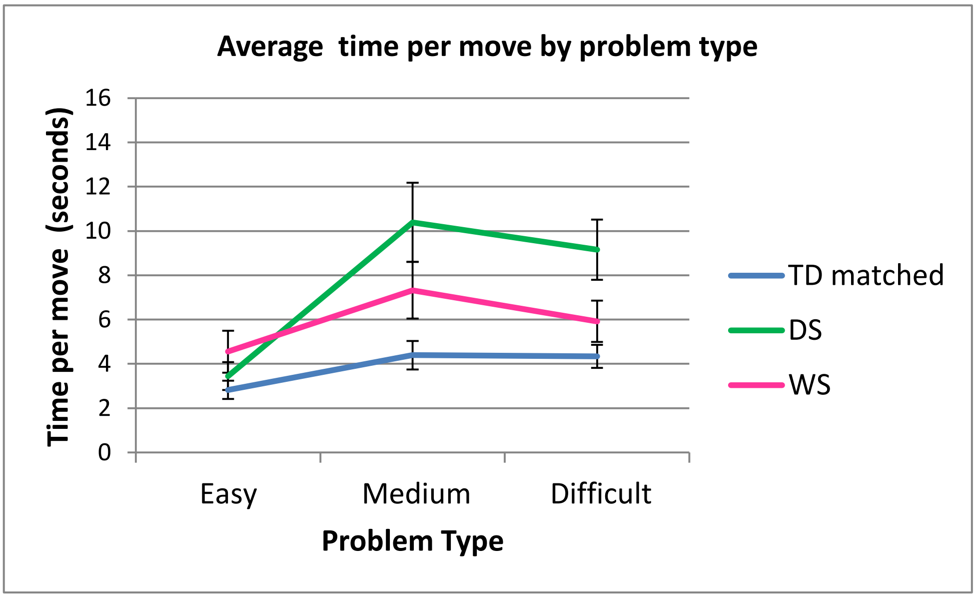

Figure 3.3: Group means *(S.E.)* of time per move (seconds) by problem type

There was a significant main effect of problem type (F(2,80) = 16.335, $p < .001$, partial $\eta^2 = .290$) with significantly longer times per move for medium and difficult problems compared to easy problems ($p < .001$ for both) and no significant difference between medium and difficult problems ($p = .231$). There was also a significant main effect of group (F(2,40) = 6.510, $p = .004$, partial $\eta^2 = .246$) with Tukey tests indicating longer times per move for the DS group than the TD group ($p = .002$) and a lack of significant differences between the WS group and either of the other groups (DS: $p = .255$; TD: $p = .134$). There was a significant interaction between Problem type and Group (F(4,80) = 3.411, $p = .013$, partial $\eta^2 = .146$), which can be explained by the presence of a significant main effect of Group for medium problems (F(2,40) = 5.399, $p = .008$, partial $\eta^2 = .213$) and difficult problems (Welch's F(2,22.809) = 5.737, $p = .010$, partial $\eta^2 = .239$) but not for easy problems ($F(2,40) = 1.632$, $p = .208$, partial $\eta^2 = .075$). The difference between DS and TD times per move was significant for medium problems (p = .006) and difficult problems (p = .003) with marginally longer times for the DS group than the WS group for difficult problems (p = .067). All other pairwise differences were non-significant ($p > .05$).

3.3.4 TOL behaviours

This section compared the behaviours that were engaged in while solving the TOL problems across the three matched groups, between correctly and incorrectly solved problems. This allows insight into what happens when a problem is failed. The number of each type of behaviour, per attempted trial, was calculated for correct and incorrect trials separately, with 1- and 2-move problems half-weighted as above.

3.3.4.1 Rule violations (RVs)

Where video data were unavailable (the 5 trials for which timing data were excluded), for the measures that were noted during testing (rule violations, reason for ending a trial) the experimenter-recorded measures were used for these participants. RV data were excluded from three trials (one TD, two DS) on which no moves were made, for which the participants were included in the matched groups. Participant numbers providing data for both correct and incorrect trials were 16 (TD), 19 (DS) and 17 (WS). Group data is displayed in Figure 3.4.

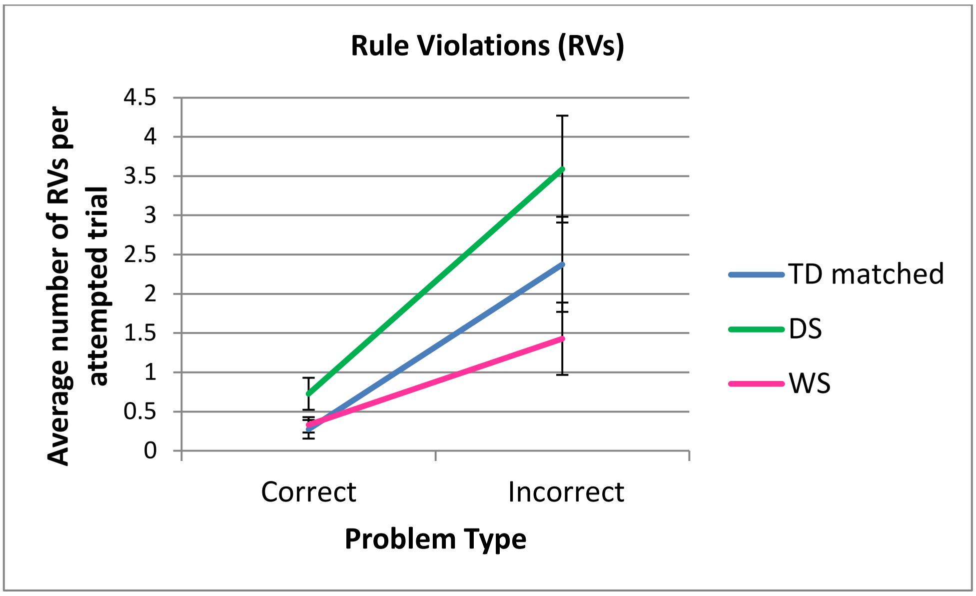

Figure 3.4: Group means *(S.E.)* of rule violations per attempted trial by accuracy

For the RVs to correct and incorrect trials, there was a significant main effect of Group ($F(2,49) = 3.642$, $p = .034$, partial $\eta^2 = .129$) with Games-Howell tests indicating more RVs in the DS group than in the WS group overall ($p = .040$) while neither atypical group differed significantly from the TD group (DS: $p = .258$; WS: $p = .531$). More RVs were made to incorrect than to correct trials ($F(1,49) = 43.228$, $p <.001$, partial $\eta^2 = .469$). The interaction between group and accuracy did not reach significance ($p = .066$).

3.3.4.2 Backup moves and incomplete moves

Backup and incomplete move data were unavailable for the three trials on which no moves were made, and for the eight trials (four from a participant with DS and four from a TD participant) for which either equipment failure or a lack of visibility meant that it was not possible to count incomplete/backup moves (while other measures were recorded by the experimenter). All of these trials were from participants in the matched groups. Thus the Ns were 15 (TD), 19 (DS) and 17 (WS). ANOVA was conducted on the backup and incomplete move data with group as the between-subjects variable (TD, DS, WS) and accuracy as the within-subjects variable (correct, incorrect). More moves of each type were made to incorrect than correct trials (backups: $F(1,48) = 37.773$, $p < .001$, partial $\eta^2 = .440$; incompletes: $F(1,48) = 50.112$, $p < .001$, partial $\eta^2 = .511$). There was no significant main effect of group for either measure (backups: F<1; incompletes: $F(2,48) = 1.218$, $p = .305$, partial $\eta^2 = .048$) and no significant accuracy by group interactions (F<1 for both).

3.3.4.3 Verbalisations

Verbalisation data were not available for the four trials from the participant with DS for whom the equipment failed, or for the three trials on which no moves were made (2 DS, 1 TD). All of these were part of the data for the matched groups. Participant numbers providing data for both correct and incorrect trials were 16 (TD), 19 (DS) and 17 (WS). The weighted average of the number of verbalisations (0 or 1) per attempted trial was calculated for each participant for correct and incorrect trials, and group means were compared using ANOVA. Again, more verbalisations were made for incorrect than correct trials ($F(1,49) = 22.966$, $p < .001$, partial $\eta^2 = .319$) while this did not interact with Group (F<1). There was a marginally significant main effect of Group ($F(2,49) = 3.157$, $p = .051$, partial $\eta^2 = .114$), which was significant using a Kruskall-Wallis nonparametric equivalent test (conducted due to non-normal TD and WS data; $\chi^2(2) = 7.859$, $p = .020$), with more verbalisations made by the DS group than the TD group ($p = .039$) and no significant differences between the WS group's number of verbalisations and those of the TD group ($p = .543$) or the DS group ($p = .348$). Group means are displayed in Figure 3.5.

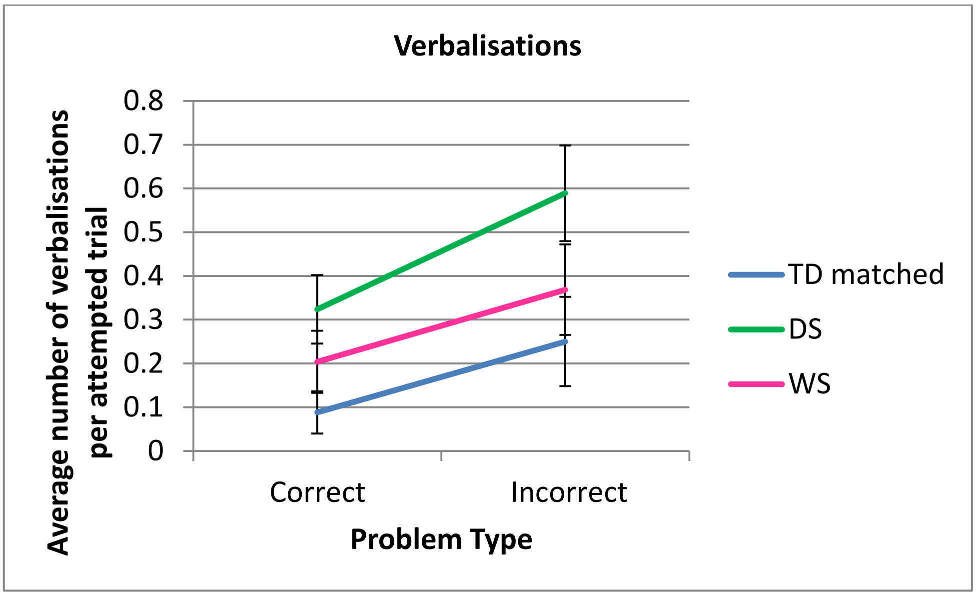

Figure 3.5: Group means *(S.E.)* of verbalisations per attempted trial by accuracy

3.3.4.4 Error types

Error types were classified as perceptual errors (PE; when the trial was stated as finished but not matching the goal state), too many moves (TMM; when the participant made 20 moves or more), or stopped (when the participant chose to end the trial because they were unable to solve it; see Table 2.2). In order to compare error types across groups, the number of errors of each type, half-weighted for 2-move problems, was calculated per participant. One participant from the WS group was not included in this analysis, because she was unable to attempt 2-move problems to the level that might admit finishing a trial and making an error, and only has scores for the 1-move trials, of which all were scored perfectly. Thus, Ns were 20 (TD), 20 (DS) and 19 (WS). ANOVA across matched groups revealed a significant main effect of error type ($F(2,112) = 5.918$, $p = .004$, partial $\eta^2 = .096$) with fewer instances of 'stopped' error types than each other type (PE: $p = .011$; TMM: $p = .001$) and no difference between the number of PE and TMM errors ($p = .525$). The main effect of Group approached significance ($F(2,56) = 3.016$, $p = .057$, partial $\eta^2 = .097$) (and reached significance using a Kruskall Wallis test: $\chi^2(2) = 6.256$, $p = .044$, conducted due to non-normal data for stopped errors in all groups, and PEs in DS and TD groups) with Tukey tests indicating marginally more errors made by the WS group than the TD group ($p = .061$) (which was significant using the Mann Whitney nonparametric equivalent, $p = .013$) with the DS group making comparable numbers of errors to the other two groups (TD: $p = .166$; WS: $p = .869$). The interaction between error type and group was not significant ($p = .197$). Descriptive data for error types by group can be seen in Figure 3.6.

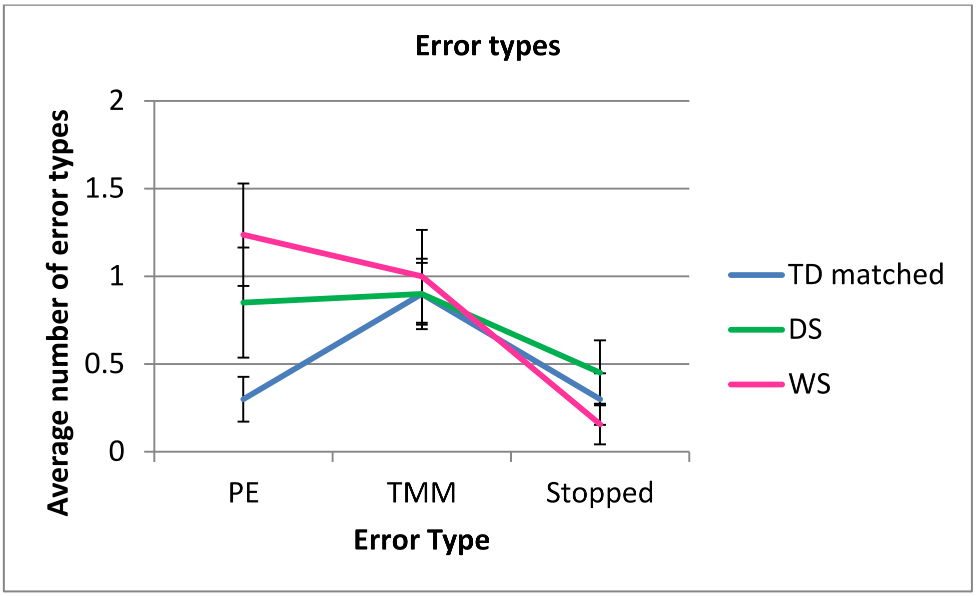

Figure 3.6: Group means *(S.E.)* of error types. PE = perceptual error, TMM = too many moves.

3.3.5 TOL correlation analysis

In order to investigate the relationships between measures on the TOL, correlations were conducted for each group between TOL measures: planning time, time per move, TOL score, RVs, incomplete moves, backup moves and verbalisations. Correlation analyses include the entire TD group, and the entire WS and DS groups, in contrast to the group comparisons for the participants who provided data for both correct and incorrect trials or for all problem types, above. This makes the values of Ns 56 (TD), 20 (DS) and 20 (WS). For this analysis, the weighted planning time; time per move; number of RVs; incomplete moves; backup moves and verbalisations per attempted trial were calculated for each participant and correlated with one another and with the TOL scores, for each group. A correlation matrix for each group is provided below. The TD group's data were treated as normal because the N was above the central limit theorem cut-off of 30. For DS and WS groups, all of the measures in Table 3.3/Table 3.4 were normally distributed; thus Pearson's correlations were conducted. Significant correlations are presented against a green background, and marginally significant correlations against a yellow background.

TD Group R (*p*)	Rule Violations	Incomplete moves	Backup moves	Planning time	Time per move	Verbalisations	TOL score	Trials correct score	Additional moves (correct trials)
CA	-.581 (<.001)	-.523 (<.001)	-.417 (.001)	-.097 (.478)	-.656 (<.001)	-.224 (.096)	-.591 (<.001)	.572 (<.001)	.136 (.319)
Rule violations		.275 (.040)	.274 (.041)	.016 (.905)	.543 (<.001)	.114 (.402)	.585 (<.001)	-.573 (<.001)	-.171 (.206)
Incomplete moves			.540 (<.001)	-.172 (.206)	.188 (.166)	.130 (.339)	.356 (.007)	-.258 (.055)	.181 (.182)
Backup moves				-.159 (.242)	.230 (.089)	.207 (.126)	.269 (.045)	-.177 (.192)	.230 (.088)
Planning time					.445 (.001)	.280 (.036)	.105 (.439)	-.220 (.103)	-.420 (.001)
Time per move						.247 (.067)	.575 (<.001)	-.588 (<.001)	-.251 (.062)
Verbalis-ations							.091 (.506)	-.106 (.435)	-.086 (.527)

Table 3.2: Correlations (r, with *p* value in parentheses) between behavioural TOL measures for the TD group. Significant correlations are presented against a green background, and marginally significant correlations against a yellow background.

The 42 correlations in Table 3.2 would require a *p* value cut-off of 0.00119 to be considered as surviving a Bonferroni correction. This applies to the correlations between CA and: rule violations, incomplete moves, backup moves, time per move, TOL score and TOL trials correct score; between rule violations and: time per move, TOL score and TOL trials correct score; between incomplete moves and backup moves; between planning time and: time per move, additional moves to correct trials; and between time per move and: TOL score, TOL trials correct score.

DS Group R (*p*)	Rule Violations	Incomplete moves	Backup moves	Planning time	Time per move	Verbalisations	TOL score	Trials correct score	Additional moves (correct trials)
CA	-.368 (.110)	-.213 (.366)	-.119 (.616)	.434 (.056)	.128 (.590)	-.084 (.726)	-.341 (.141)	.310 (.183)	-.042 (.861)
Rule violations		.574 (.008)	.161 (.497)	-.078 (.744)	.092 (.700)	.171 (.472)	.212 (.369)	-.090 (.705)	.545 (.013)
Incomplete moves			.725 (<.001)	-.053 (.825)	-.177 (.454)	.100 (.675)	.105 (.659)	.033 (.890)	.624 (.003)
Backup moves				-.142 (.551)	-.357 (.122)	.015 (.949)	.151 (.525)	-.066 (.783)	.331 (.155)
Planning time					.662 (.001)	.176 (.459)	-.175 (.462)	.118 (.621)	-.096 (.686)
Time per move						.369 (.110)	.172 (.468)	-.211 (.373)	-.150 (.529)
Verbalis-ations							.262 (.265)	-.199 (.401)	.253 (.282)

Table 3.3: Correlations (r, with *p* value in parentheses) between behavioural TOL measures for the DS group. Significant correlations are presented against a green background, and marginally significant correlations against a yellow background.

For the DS group, the correlations surviving a Bonferroni correction ($p \leq 0.00119$) would be between incomplete moves and backup moves; and between planning time and time per move only.

WS Group R (p)	Rule Violations	Incomplete moves	Backup moves	Planning time	Time per move	Verbalisations	TOL score	Trials correct score	Additional moves (correct trials)
CA	.076 (.750)	-.288 (.218)	-.491 (.028)	.320 (.169)	.357 (.122)	.230 (.329)	-.187 (.429)	.162 (.495)	.012 (.959)
Rule violations		.739 (<.001)	.043 (.856)	.178 (.452)	.704 (.001)	-.052 (.826)	.208 (.378)	-.148 (.533)	.501 (.025)
Incomplete moves			.321 (.167)	.006 (.980)	.336 (.148)	-.063 (.792)	.243 (.302)	-.207 (.381)	.259 (.270)
Backup moves				-.274 (.243)	-.137 (.565)	-.171 (.470)	-.080 (.739)	.111 (.641)	.246 (.295)
Planning time					.611 (.004)	.410 (.073)	-.397 (.083)	.391 (.088)	.170 (.473)
Time per move						.086 (.718)	-.082 (.730)	.133 (.576)	.494 (.027)
Verbalis-ations							-.201 (.395)	.168 (.480)	-.164 (.490)

Table 3.4: Correlations (r, with p value in parentheses) between behavioural TOL measures for the WS group. Significant correlations are presented against a green background, and marginally significant correlations against a yellow background.

For the WS group, only the correlations between rule violations and incomplete moves, and between rule violations and time per move, would survive a Bonferroni correction at $p \leq 0.00119$.

3.4 Discussion

The aim of this chapter was to explore behaviour on the TOL task in TD, DS and WS groups, in order to understand the processes involved in their performance. Each element of task performance will be discussed in turn.

3.4.1 Problem type

Three problem types were identified within the problem set: easy (2-3 moves), medium (4 moves) and difficult (5-6 moves). When examining the problem set used on the TOL task (Appendix A), we can see that while 3-move problems require more moves than 2-move problems, by design they possess the same search depth and number of counterintuitive moves as 2-move problems. The lack of significant difference between the ability to solve the two types of problems for the matched groups supports previous assertions that the minimum number of moves should not be the only indicator of problem difficulty (Kaller et al., 2004; Berg et al., 2010). Indeed, Kaller et al. (2008) noted that the presence or otherwise of a counterintuitive move affected the success that 4-year-olds had with solving 3-move problems. In encountering the medium problems in the present study (of 4 moves), this was the first time that solutions required a search depth (and number of counterintuitive moves) greater than zero, and success duly declined, although an additional move was also required to reach the solution for the 4-move problems compared to the 3-move problems, highlighting the often intertwined nature of the parameters concerned. Interestingly, between 5- and 6-move problems, success rate did not differ, while the minimum number of moves, search depth and number of intermediate and counterintuitive moves did differ. That is, while the first subgoal in the 6-move problems required 3 intermediate moves to reach it, the first subgoal in the 5-move problems required 2 intermediate moves. This difference does not appear to have affected performance. However, in order to keep other parameters constant it was necessary to select 6-move problems that had 5 detour paths, that is, 5 non-optimal ways of reaching the solution, while the other problem types had 1 detour path (and the number of optimal paths was 1 for every problem). Thus it is more likely that a correct solution could have been found by chance in the 6-move problems than in the 5-move problems. Importantly, without knowledge of the properties of each item in the problem set, one would need to draw conclusions based entirely on the minimum number of moves to the solution, and an understanding of the potential reasons for differences (or lack thereof) in scores would not be known. This highlights the importance of providing information about the parameters of the problem set that is used to facilitate comparison between studies (cf. Berg & Byrd, 2002).

Rhodes and colleagues' (2010) WS group made more moves than at least one control group on 3-, 4- and 5-move problems, while extra moves for 2-move problems did not differ. In contrast, in the current study there were similar patterns of performance for each group with regard to problem type. This could be due to differences in the problem set: the CANTAB uses Shallice's (1982) original set of 12 problems, in which one of the 2-move items is flat-ending, while the current study uses partial tower-ending problems. Flat-ending trials are more difficult than partial tower-ending problems (Spitz et al., 1982), and thus they could have elicited more moves from the WS group in Rhodes et al. (2010)'s study.

The pattern of correlations between EF and TOL scores for easy, medium and difficult problems indicates differences across groups in the way that the strength of the association varied for different levels of task difficulty. For the TD group, the significance of the association generally increased as problems become more difficult: e.g., some EF tasks only showed a significant relationship for the most difficult problems (both backwards memory tasks, forwards digit span) while others showed relationships for medium and hard problems, but not easy problems (forwards block span, RT Same and Opposite, planning and BPVS). Even for the RCPM which was significantly related to all problem types, the R value increased with each level of task difficulty. This pattern of outcomes supports the general concept of EFs becoming more important for tasks which require more deliberate, effortful control (Rabbitt, 1997). This fits with the idea that when beginning an easy (2- and 3-move) problem it may be that the route to solving the problem is apparent from the start, so that deliberate memory, planning or shifting capabilities are not needed: then, only when the task becomes a challenge do these abilities, tapped by the EF tasks, come into play for problem solving. This confirms the earlier-stated expectation that there would be more contribution of EF skills to more difficult problems than to easier problems.

This pattern extends to the WS group's association between TOL scores and both backwards block span and planning scores, with non-significant associations with

TOL score for easy problems that increased in strength for medium and difficult problems. The greater reliance, in general, on backwards visuospatial memory and planning abilities in the WS group on more difficult problems is consistent with the idea that the more difficult problems require more manipulation of items in working memory or visualisation abilities. The reason for this seems clear when we consider the problem set. For 2- and 3-move problems, each move is a goal move, so the planning ahead/manipulation requirements, while present, are quite low. However 4-move problems require one temporary move away from the goal state to solve efficiently, while harder problems demand this type of ability even more. On some TOL problems (but not 4-move problems here) temporary moves require the participant to place a piece temporarily in a different piece's goal state, or temporarily remove a piece from its own goal state, which could be difficult for children, being related to Klahr (1985)'s concept of 'sensitivity to partial evaluation'. These types of moves are likely to require more planning ahead and manipulation of items in memory to solve efficiently (e.g., Kaller et al., 2008). The relationship between these skills and TOL performance in WS as problems get harder adds to the suggestion in Chapter 2 that these may be limiting factors for spatial problem solving in WS (although planning may only be a limiting factor at a low level of planning ability).

Interestingly, backwards digit span and RCPM only showed associations (marginally for RCPM) to medium difficulty problems in the WS group. The relative difficulty of the 4-move problems could account for this: for example, reasoning abilities (RCPM) could be used to generate a new strategy for medium problems, which could then be adopted (or adapted) for difficult problems, without demanding reasoning abilities to generate another new strategy. The pattern of the longest planning times to medium difficulty problems, followed by a drop in planning times to difficult problems, supports this explanation.

In contrast, for some measures the pattern of associations in the WS group was in the reverse direction: the association between BPVS and TOL score decreased in strength with increasing TOL difficulty, while accuracy on the inhibition task was only related to the TOL score for easy and medium problems. Shifting and forwards digit span were also marginally related to the TOL for the easy problems alone. Similarly, in the DS group, the backwards digit span association was stronger in easy than medium problems and unrelated in more difficult problems. In fact, the only measure related (albeit marginally) to TOL score for difficult problems in this group was RT Same.

It is possible to speculate about the root of the pattern of stronger relationships with easier problems in the atypical groups. It may be that easier problems, being within a participant's general level of capability, are readily solvable as long as they are attended to. As problems become more difficult, the increasing reliance on some EFs in the WS group but decreased reliance on EFs in the DS group may reflect a discrepancy in the groups' problem-solving processes: the WS group may have engaged in planning/mental manipulation (which was difficult for them) while the DS group may not have engaged in processes measured by the EF tasks (potentially, attempting rule violations, addressed later in the section). Indeed, it was suggested in Chapter 2 that the DS group were able to compensate for poor planning abilities, as they were poorer than the TD group on the planning task but just as good on the TOL. Choosing to use an alternative strategy could be an attempt to compensate for poor planning abilities (but see Section 3.4.4 below for further discussion of rule violations in the DS group). An alternative reason for a scarcity of relationships between some EF tasks and difficult TOL problems, particularly in the DS group, might be a strategy of moving the pieces at random until the correct solution, or a near solution, is chanced upon.

3.4.2 Error types

The marginally significant difference in the number of errors made between WS and TD groups is consistent with the marginally lower TOL trials correct score in the WS group than in the TD group seen in Chapter 2. While most participants attempted to complete the trials rather than producing a 'stopped' error, there were no other differences. Thus, the groups did not significantly differ in their reasons for failing a trial.

3.4.3 Timing

3.4.3.1 Problem type and group comparisons

Planning time was longest for medium problems, and shortest for easy problems, with planning time for difficult problems being longer than easy problems and shorter than medium problems. As medium problems were the first instance of a problem that requires a temporary, and moreover a backwards temporary, move in order to solve them, it seems likely that these longer planning times were (at least partly) due to this requirement. It is possible that these were the first problems for which the solution was not immediately apparent and thus required more time to generate the first move. Four-move problems were also the first problem type in which the start and goal states consisted of configurations of pieces on the same two pegs, perhaps making it less apparent what the first steps should be, that is, how to rearrange the pieces between the two pegs (see Appendix A). This suggests that a certain amount of planning took place before any moves were made. Shorter planning times for difficult than medium problems suggest that the greatest demands on preplanning were not for the most difficult problems *per se*, but for the problems that posed the greatest cognitive challenge with respect to what has previously been encountered (i.e., medium problems).

The increase in planning times from easy to medium problems is consistent with previous research with adults, who displayed longer planning times to problems that were more difficult (Unterrainer et al., 2004) and had more indirect (non-goal) moves (Phillips, 1999). These problem sets, however, being designed for use with adults, were more challenging than the set used in the current study, thus limiting the value of such comparisons. The decrease in planning times from medium to difficult problems here is not consistent with these previous findings. It may be that the pattern of longer planning times to harder problems is confined to adults; indeed, in Unterrainer et al. (2004)'s study the pattern of longer planning times to harder problems was most pronounced for the better planners. The discrepancy between the results for adults and for children and individuals with neurodevelopmental disorders found here points to the need to avoid assuming that patterns of performance seen in adults will extend downwards to children (cf. Karmiloff-Smith, 1997). Alternatively it may be that the relative novelty of the 4-move problems increased planning times to such an extent as to mask a gradual increase in planning time with task difficulty.

Planning times were marginally longer in the DS group compared to a TD group of the same nonverbal ability. This is reminiscent of the slightly longer RTs seen on the inhibition task by the DS group in Chapter 2. RT can be understood to include both cognitive processing and motor processing. A measure of motor processing speed is not available in the current study, and would be a useful inclusion in further work.

For the WS group, planning times did not differ significantly from those of RCPM-matched controls, while in Rhodes et al. (2010)'s study they were longer in a WS group and a CA-matched group compared to verbal mental age-matched group, who were younger than the other two groups. Planning times were longer in the WS group than the other two groups for 2- and 3-move problems, while the CA-matched group's were longer than the VMA matched group's for 4- and 5-move problems; the WS group's planning times were equivalent to those of each control group.

Time per move did not differ between the WS and TD groups in the current study. Again, seemingly at odds with this, Rhodes et al. (2010) found longer subsequent thinking times in their WS group compared to at least one of the control groups, for each type of problem. This could be due to administration differences: Rhodes and colleagues were able to control for movement time by using a yoked control procedure, while the current measure includes movement time as well as thinking time. Also, Rhodes and colleagues' WS group made more moves than at least one control group on three out of the four problem types, so it is likely that subsequent thinking time was confounded by the number of moves made, while in the current study this was controlled for by calculating the time per move. Finally, group matching measures were different: poorer performance (if longer times reflect poorer performance) compared to a verbal or CA-matched group, but not compared to a nonverbal-matched group, can be expected given the uneven cognitive profile in WS (Mervis et al., 2000). Indeed, other studies comparing time-based measures on the TOL found equivalent total time scores in WS groups compared to nonverbal matched controls (Costanzo et al., 2013; Menghini et al., 2010). While any comparisons to the current study cannot be wholly accurate because of measurement differences, as in the present study neither the planning time, time per move nor number of additional moves (to correct trials) differed from that of the control group, these findings seem to be consistent with those found previously.

The DS group displayed significantly slower times per move than the TD group for medium and difficult problems (and slightly longer times than the WS group for difficult problems). The finding of longer TOL times, in general, for participants with DS compared to MA controls is consistent with previous studies who found longer total solution times in a DS group (Costanzo et al., 2013; Rowe et al., 2006) and consistent with the slower processing expected for the DS group. Present results extend previous findings because timing measures were split into planning time and execution time (per move): as it is the time per move during solving for which this difference reached significance and not the planning time, this could either imply slower motor processing by the DS group during the task (which could not affect planning time as much) or more thinking 'on-line' during the task in the DS group than the TD group, or both.

Times per move were longer in medium and difficult problems than in easy problems. This might reflect a greater amount of on-line planning in problems that are more demanding. The difficult problems elicited less planning time, but an equivalent length of execution time per move, than the medium problems. This seems to add weight to the idea that the planning time increased so dramatically from easy to medium problems because of the unexpected configuration of the pieces: the difficult problems would have been less likely to have elicited this unexpectedness after encountering the medium problems, and indeed, planning times were not as long.

If participants were planning all of their moves prior to beginning a trial, we would have seen increasing planning time with problem difficulty, but equivalent times per move. This was not the case. This is in line with previous suggestions that a perceptual strategy (i.e., planning on-line) is the dominant one when participants are not given strategy-specific instructions (Goel & Grafman, 1995), but also with suggestions that children do not simply engage in trial and error when problem solving (Klahr & Robinson, 1981); rather, they apply "general problem-solving methods" (p. 144).

3.4.3.2 Correlations

Several correlations were conducted between performance measures on the TOL in order to understand performance. To add a note of caution, conducting a large number of correlations runs the risk of producing type 1 errors. It was expected that longer planning time would be associated with better TOL performance if participants were using that time to plan. In the TD group, longer planning time was related to making fewer additional moves to correct trials, suggesting that children may indeed have been engaging in useful planning on some of the trials (as suggested above). This association was not present in either of the atypical groups, indicating that if individuals with WS or DS were actively planning, this did not produce more efficient correct solutions. It is equally possible that planning time did not reflect active planning for the atypical groups, or that their plans were not seen through to fruition, perhaps because of working memory difficulties.

However, longer planning times were marginally related to better TOL accuracy in the WS group, with no relationship in the DS or TD group. This, tentatively, suggests that the avoidance of impulsive responding by taking longer to plan may have been beneficial for performance in the WS group. In the TD group, longer times per move were also related to poorer TOL accuracy, perhaps representing trials for which they were unsure of the solution and were engaging in a large amount of planning on-line (i.e., during solving). There was also a (marginal) association for the TD group in the opposite direction: between taking more time per move and making fewer additional moves in correct trials. Thus, TD participants may be engaging in more on-line planning when problems are difficult, but when they were within their capability, the longer move time could have improved efficiency. The marginal nature of the latter correlation makes this a tentative suggestion.

For the WS group, longer times per move were associated with more additional moves to correct trials, that is, poorer efficiency. Thus, for the WS group, taking longer to plan was associated with increased success in some form (better accuracy), while taking longer to move was associated with reduced success in some form (poorer efficiency). Unterrainer et al. (2004) found that, in typical adults, shorter execution times and longer planning times were associated with better accuracy on the TOL. For the TD group, this was found in relation to shorter move times but not for longer planning times. For the DS group, the only TOL measure that longer times per move was associated with was longer planning time. This relationship was also seen for the other two groups in the same direction. It might be that individual participants either have an early insight into the strategy to take, making both planning time and time per move short, or else have difficulty throughout, taking a slow and deliberate approach to both planning and making subsequent moves. Alternatively, it may be that an individual's processing speed dictates the time taken in both planning and solving phases, rather than more time taken planning facilitating faster subsequent moves (as is implied by Unterrainer et al., 2004's finding, above, of both increased planning time and reduced execution time being related to better TOL performance, although they do not report whether or not there is a direct association between the two timing measures). As Unterrainer et al. (2004)'s findings are with respect to typical adults, processing speed may be the overriding factor determining planning and execution times for children and individuals with neurodevelopmental disorders, while it may be that the pattern of longer planning and short execution time facilitating performance only applies to adults (i.e., better planners). Indeed, in the TD group, the association between longer times per move and poorer TOL accuracy (poorer trials correct scores), which was suggested above as representing difficult trials, may alternatively be due to differences in processing speed with age. As we know that processing speed (RT Same) increased with age, and TOL scores also increased with age (see Section 2.3.5.1), perhaps the association between longer execution times (i.e., slower processing) and poorer TOL scores is produced because the younger TD children, being slower processors, also scored more poorly on the TOL.

3.4.4 Rule violations (RVs)

3.4.4.1 Group comparisons

There were more RVs made overall on incorrect trials than correct trials. In addition, the DS group made more RVs than the WS group. In contrast to the WS group, for whom associations between planning and backwards block span scores and the TOL score became stronger with increasing problem difficulty, none of the EFs displayed this pattern in the DS group. This suggests that when trials were difficult (and thus, more likely to be solved incorrectly), rather than attempting to engage with the more difficult problems by planning ahead or using working memory, the DS group were attempting to move the pieces to the correct position by making moves that are illegal in the task.

Berg and Byrd (2002) noted that RVs could occur either because of rule forgetting or because of choosing not to follow the rules, and they suggest that deliberately not following them may be more likely in typical populations. Dual-task processing, that is, performing two tasks simultaneously, is impaired in DS (Lanfranchi, Baddeley, Gathercole, & Vianello, 2012) and disproportionately more difficult than performing one task in DS than in WS (Kittler, Krinsky-McHale, Devenny, & Conners, 2008). In the TOL, one needs to remember the rules of the game in addition to moving the pieces in order to solve the problem. It seems possible that a higher rate of RVs in the DS group than in the WS group could be related to greater difficulties with dual-task processing in DS than in WS. Thus, the greater number of RVs in the DS group could either be due to difficulties with the task, or a deliberate strategy.

3.4.4.2 Correlations

In the TD group, making more RVs was related to lower levels of accuracy on the TOL, while for the atypical groups it was related to making more additional moves to correct trials, that is, to less efficient solving. Neither of these outcomes are surprising, as we know that more RVs occurred on incorrect than correct trials, and that greater numbers of RVs will necessarily increase the numbers of additional moves recorded.

3.4.5 Further measures

3.4.5.1 Verbalisations

The presence or absence of a task-relevant verbalisation made by the participant during problem solving was recorded for each trial. Self-directed speech includes overt (audible) private speech and covert (inaudible) inner speech (cf., e.g., Lidstone et al., 2010). Fernyhough and Fradley (2005) found that TOL performance was associated with private speech in 5-6 year old TD children, although the relationship between private speech and task difficulty is often not found to be linear but, rather, more private speech is made to problems that are of medium difficulty (e.g., Fernyhough & Fradley, 2005). Fernyhough noted that this is consistent with the Vygotskyan idea that when a task is within the zone of proximal development, that is, challenging but still within a child's capability level, this is when most private speech occurs.

In the current study, more verbalisations were made to incorrect trials than to correct trials overall. However, making more verbalisations was not related to TOL success in any of the groups; thus, it does not seem to have facilitated performance. It is unclear which tasks might have been in each participant's zone of proximal development (thus eliciting more private speech), and of course, it is not possible to say whether or not inner speech was associated with performance, as we are only able to measure audible speech. In fact, if self-directed speech becomes more covert with development, this could explain a lack of association between audible self-directed speech and TOL success, as older children would be likely to both make fewer verbalisations and to score better on the TOL. This is supported to some extent by the marginal negative correlation between age and verbalisations for the TD group.

It was expected that more verbalisations would be made by the WS group, as they have been previously suggested to use verbal mediation strategies. However, it was the DS group who exhibited more verbalisations than the TD group. For the TD and WS groups only, increased planning time was related to making more verbalisations (marginally so for the WS group). This suggests the occurrence of verbal planning for these two groups, but not in the DS group.

Lidstone et al. (2010) found that articulatory suppression (which removes the ability to produce self-directed speech) disrupted children's TOL performance when they were required to plan their moves prior to solving, but not when they were merely encouraged to think ahead. They suggest a link between planning and self-directed speech. As in the present study no demand to plan beforehand was made, this could explain the lack of association between verbalisations and TOL success. However, perhaps, when planning does happen, it is reflected verbally to some extent in the TD and WS groups; hence, the links between planning time and verbalisations for these groups. Given the lack of association between planning time and verbalisations in the DS group, it may be that the DS verbalisations, of which there were more than in the TD group, may have been predominantly other types of verbalisation, such as saying out loud what is difficult about the task or giving self-encouragement. This is still self-directed, but may not be strictly classed as planning. As verbalisations were recorded at a binary level for each trial, potential analysis of the exact type of verbalisation or the duration of the verbalisation per trial would be a potential avenue for future research.

3.4.5.2 Incomplete and backup moves

More incomplete and backup moves were made to incorrect trials than to correct trials, and there were no significant group differences for either measure. In the TD and DS groups, incomplete and backup moves were correlated with one another, which might then reflect on-line planning indicated by both behaviours. For the TD group there was a relationship between more incomplete moves and poorer TOL accuracy, but this could stem from the extra moves made to 'too many moves' error trials. Incomplete moves (in DS) and backup moves (marginally, in TD) were associated with making more additional moves to correct trials, which again is unsurprising.

3.4.6 Remaining correlations between measures

It was expected that incomplete/backup moves would be associated with greater times per move if they reflected on-line planning. Backup moves, and verbalisations, were indeed (marginally) related to greater times per move for the TD children, but this was not seen for incomplete moves. This could reflect the additional time needed to make two moves (a backup) and to make verbalisations, rather than reflecting the on-line deliberation associated with this type of move.

Greater times per move were also related to making more RVs for the TD and WS groups (but not for the DS group) which is as to be expected because RVs will be likely to take more time per move as the experimenter must reset the pieces. Making more RVs was related to making more incomplete moves for all three groups, and backup moves were related to making more incomplete moves for the DS group, and more RVs for the TD group. This might reflect challenges experienced during the trial that manifest in making fewer straightforward peg-to-peg moves, and more moves of other types.

3.4.7 Correlations with CA

In the DS group, planning time increased with age, although this did not reach significance, while in the WS group, the number of backup moves reduced with increasing CA. Although these measures appear to indicate more efficient performance with age (presuming that planning time does reflect more planning), notably, neither of these measures were related to TOL success for the group concerned, and neither atypical groups' TOL performance increased with age.

With increasing CA, the TD group made fewer errors of all three types, fewer backups, fewer incomplete moves, fewer RVs and took less time per move. Verbalisations also marginally decreased with increasing age, possibly reflecting the age-related shift from private speech to inner speech discussed above. Thus, with increasing age, performance became more accurate, with fewer indications of on-line deliberation during solving. Of course, it is possible that deliberation during solving was still occurring for the older TD children but that it was not manifested in these behavioural measures. Indeed, this pattern might reflect increases in processing speed with age (correlation between RT Same and CA) that mean that on-line planning is achieved without the need for incomplete moves and backups, or increased times per move. Planning time did not change with CA in the TD sample. This is consistent with the findings of Culbertson and Zillmer (1998).

The lack of age-planning time association could either be interpreted as representing the same amount of pre-planning and on-line planning as age increases but with more efficient-looking 'in-trial' performance with increasing age-related processing speed, or as planning times that reflect different processes for younger and older TD children. As noted above, planning time might reflect cognitive difficulty or active planning (Berg & Byrd, 2002). If, for example, planning time for younger children reflects cognitive difficulty but for older children it reflects *bona fide* planning, this shift would not be visible in the constant relationship between planning time and age. In addition, in the current study older children are likely to have attempted more difficult problems than younger children, and planning times were longer for medium and difficult problems than easy problems (although longest for medium problems). If all participants had completed all the trials, the relationship between planning time and age would have been easier to interpret.

3.4.8 Chapter summary

In TD children, performance on the TOL became more efficient with age, while this was not the case for either of the atypical groups. TD children showed associations of greater strength between EFs and TOL problem solving with increasing problem difficulty, suggesting that they were engaging more effortful executive processes in more difficult trials. Longer planning time was also associated with better TOL performance, and more verbalisations. These measures combine to form a picture of children being more involved in more difficult problems, and increasing their planning time and verbalisations to help them to solve problems.

Behaviour looked quite different for the DS group: TOL performance was not associated with increasing age; generally greater EF associations were seen for easier than more difficult problems; the DS group took slightly longer times per move than the TD controls but did not show the same associations between that execution time and their TOL performance, suggesting that they were taking longer for different reasons. The DS group also engaged in more verbalisations, but unlike the TD group, these were not related to increased planning time, again potentially indicating the occurrence of different underlying processes.

The WS group appeared to display some typical-looking patterns of performance, for example, greater reliance on some EFs with increasing problem difficulty, and an association between longer planning times and more verbalisations. In contrast, some patterns were atypical, with some EFs showing greater TOL associations to easier problems, and with an association between time per move and solving efficiency that was in the opposite direction to the (marginal) TD association.

The combination of greater numbers of RVs in the DS than the WS group, and decreased reliance on EFs with increasing problem difficulty in the DS group, suggests syndrome-specific differences in problem-solving approach between the two atypical groups: the DS group may have been making illegal moves rather than engaging in a deliberate way to figure out a difficult problem.

CHAPTER 4: EVERYDAY MEASURES OF EXECUTIVE FUNCTIONS AND PROBLEM SOLVING

4.1 Introduction

In this chapter, a parental questionnaire study is reported in which the aim was to investigate and understand problem-solving abilities in everyday life in TD, DS and WS groups. The purpose of using questionnaires was two-fold: first, to understand the relative profiles of everyday problem solving in order to help point the way towards interventions for these populations and second, to enable assessment of the ecological validity of the experimental EF tasks that were used earlier in this thesis, for individuals with WS and DS.

4.1.1 Problem solving inside and outside the laboratory

Problem solving has traditionally been studied in a laboratory environment, following neuropsychological and computer science background influences, as discussed in Chapter 1. Problem solving actually happens, most of the time, not in a controlled, artificial testing situation but in the multifaceted, unpredictable and complex everyday world. Problem-solving skills are drawn on when, for example, catching a bus, tying a shoelace, making a cup of tea, and so on. Problem solving, of course, is not only carried out by individuals, but frequently in collaboration with others. Social problem solving, e.g., how to resolve conflict with others, is another sphere of research that is beyond the scope of this thesis, but nevertheless just as applicable to everyday life.

Imagine, if you will, that a person with a neurodevelopmental disorder has been to visit a university to take part in a research study, along with an accompanying parent or carer. During the study they will likely have been shown where to sit, given all the materials and equipment they need, been given very specific task instructions, their understanding will have been checked, and if they forget along the way what they are to do, it is likely that they will readily be reminded. When they finish the task, they will be praised, and likely rewarded; the testing session is over. Now imagine that they leave the university campus: what to do next?

How to get home? Should they go straight home, or visit the shops first? Does this depend on the weather, the time of day, or the season? They are once again operating in the multifaceted, unpredictable and complex everyday world. This simple juxtaposition of demands on the individual speaks to two related issues in the literature with respect to problem solving both in, and out, of the laboratory: i) the ecological validity problem, and ii) well- and ill-defined problems in problem-solving research.

Both of these issues are discussed in Chapter 1. Briefly, the ecological validity of EF tests has been challenged (Burgess et al., 2006): to what extent does what we measure on an experimental task represent, or generalise to, what happens in the real world? Although some links have been found between EF tests and measures of everyday functioning, the utility of asking about EF in everyday life has also been recognised (Burgess, Alderman, Evans, Emslie, & Wilson, 1998). Ecological validity is receiving increasing attention in the area of neuropsychology (Spooner & Pachana, 2006), with several researchers investigating the links between EF measurements in experimental and everyday settings, often with reference to neuropsychological patient groups (Burgess et al., 1998; Chaytor, Schmitter-Edgecombe, & Burr, 2006; Davies, Field, Andersen, & Pestell, 2011; Odhuba, Broek, & Johns, 2005) as well as investigating the predictive value of EF tests for everyday life, e.g., the WCST for predicting abilities relevant to working life for adults with head injuries (Kibby, Schmitter-Edgecombe, & Long, 1998) and the block design test for predicting everyday spatial ability in typical adults (Groth-Marnat & Teal, 2000). More recent research also highlights the importance of investigating EF in real life, because it has been suggested that test performance and real-life rating scales do not measure the same underlying construct (Toplak, West, & Stanovich, 2013).

The distinction between well-defined and ill-defined problems is also of relevance. The TOL can be classified as well defined, because the start, end, means of reaching the end, and rules are clearly defined (Kahney, 1986a). Well-defined problems form the basis of most problem-solving research studies (Dunbar,

1998) while psychology has made little progress in understanding ill-defined problems (Pretz et al., 2003) which are much less confined and controlled and, almost by definition, more likely to be encountered in everyday life. The goal or start state may not be clearly identified; there may be many different factors to take into account; there may also be several correct solutions; moreover, the problem itself may be difficult to identify, and the operators unclear (Kahney, 1986a; Pretz et al., 2003). One example of an ill-defined problem is arranging to meet a friend: at the outset it is not clear which friend to meet, where to meet, for how long, how and when to contact them, and so on. It is perhaps of little wonder that more research progress has been made in relation to well-defined than ill-defined problems. If people are to be supported in daily problem solving, it is imperative that efforts are made to investigate ill-defined problem-solving processes; particularly the "fuzzy issues surrounding problem recognition, definition, and representation" relevant to ill-defined everyday problems (Pretz et al., 2003 p.27).

4.1.2 Everyday life in WS and DS

We will now consider what is already known about everyday functioning in WS and DS. While the following studies do not address problem solving or executive functioning *per se*, they do address adaptive functioning and independence, which have a relevance to problem solving.

4.1.2.1 DS

In a longitudinal interview study, Carr (2008) reports the most recent outcomes regarding everyday life for people with DS, from 15 months old, followed by interviews at 4, 11, 21, 30, 35 and 40 years (34 participants remained in the study by 40 years of age). Parents of individuals with DS were engaged in semi-structured interviews at each time point. At age 40, 77% of individuals attended a social education centre either full or part time; 36% a further education college, part time; and 21% were in part time work (0% in full time education or work). Independence measures were collected addressing whether the individual was left alone in the house for more than 30 minutes or for more than 60 minutes and whether they were allowed out by themselves beyond the home and garden.

Independence measures were reported to have remained generally constant from age 21 to age 40 (at all the time points in between). At age 40, 69% were not left alone in the house for more than 60 minutes (31%, not more than 30 minutes) and 53% were not allowed out, beyond the garden, by themselves. Greater independence was associated with greater nonverbal IQ (and going out, to a lesser extent, with verbal IQ also), and was related to IQ throughout the longitudinal study (Carr, 2008). Of the cohort, 11 had become lost, and for some this involved incorrect public transport journeys with police returning them home.

Carr (2012) reported on the longitudinal study as a whole. In general, tests measuring verbal, nonverbal and self-help skills remained constant over developmental time, although the onset of dementia was seen in some individuals. Nonverbal ability scores increased between ages 11 and 21 and then remained level after this. Self-help skills increased from age 11 to age 30, then declined somewhat, but not as far as the initial mean at age 11, although these seem to have been measured using different scales at age 11 compared to age 21 and above (from age 21 upwards, the MRC Handicaps, Behaviour and Skills schedule (Wing, 1980) is reported to have been used but it is not made explicit that this is the measure used for self-help scores). Feeding, washing and toileting are reported as the items responsible for a peak at age 30, although insufficient detail of the self-help measure used is provided.

Pennington et al. (2003) gave parents of children with DS aged 11-19 years the Scales of Independent Behaviour-Revised (SIB-R; Bruininks, Woodcock, Weatherman, & Hill, 1996). Participants were matched on MA to controls on the school age version of the Differential Abilities Scale (DAS; Elliott, 1990). Raw scores indicated that the DS group scored better than MA controls on Social/Communication, Personal Living Skills and Community Living Skills, but not differently on Motor Skills. Pennington and colleagues note that the difference in CA between the MA-matched groups was large (the DS group were approximately 10 years older) so they had the opportunity to gain more adaptive behaviour skills over time than their 5-year-old counterparts. However, the DS group's scores were poorer than the level that would be expected for their CA, with 79% meeting the criteria for an indication of learning disability. There were no score differences between younger and older DS groups. Adaptive behaviour was best predicted by age in the DS group, rather than skills attributed to hippocampal or prefrontal brain areas.

4.1.2.2 WS

For individuals with WS, independence and daily living skills are poor, with 53% of a sample of adults aged between 19 and 55 years living at home with their parents and a further 39% in residential care (Elison, Stinton, & Howlin, 2010). Low numbers were engaged in typical or supported employment (15%), with 21% undertaking voluntary work, mainly with extra supervision, 53% in some form of education or training, and 11% not engaging in this kind of activity. Over time, the number of adults living at home dropped significantly (Elison et al., 2010).

In the cross-sectional cohort there were no significant differences between three age groups (19-29, 30-39, 40+ years). As a whole group, in semi-structured interviews Elison et al. (2010)'s results of self-care and independence questions varied across the group, and did not change across age groups, with 77% of adults obtaining the best response (performing self-care tasks with little or no help) and 59% not needing reminders for self-care, with much smaller percentages responding at the opposite end of the scale. Most (59%) could not carry out household chores, with 16% reported as able to carry out most household tasks. Regarding independent travel, 38% were able to do this at least locally while 49% engaged in no independent travel at all. Thirty five percent needed supervision for most or all of the time. Social activities were organised at least most of the time by 80% of the sample, while 32% and 49% did not understand how to manage finances and could not make their own purchases, respectively. In a longitudinal study (drawing on data from earlier work by Davies, Howlin, & Udwin, 1997; Howlin, Davies, & Udwin, 1998), however, self-help skills from the interviews had improved over 11-12 years. Adaptive behaviour in general measured by the Vineland Adaptive Behaviour Scales (VABS) had also improved over 12 years, with improvements on component scores for Daily Living Skills and Socialization and a reduction in Maladaptive behaviours (Howlin et al., 2010), although it is possible that this at least partly reflects a change in the way that society views individuals with developmental disorders. Full-scale IQ (FSIQ), verbal IQ (VIQ) and performance IQ (PIQ) was correlated with the overall VABS score, while Daily Living skills, as well as Communication and Socialization, were correlated with FSIQ.

Rhodes et al. (2010) collected both experimental and parental rating data. They administered the CANTAB battery of EF tasks as well as non-executive memory tasks to a group of adolescents and adults with WS as well as CA- and verbal MA-matched control groups. Parents of adolescents with WS completed the Connors Rating Scale, measuring behaviours associated with ADHD, and the Strengths and Difficulties Questionnaire, giving scores for various social and emotional aspects of behaviour, e.g., 'prosocial behaviour', and 'conduct problems'. Abnormal or borderline scores were observed in between 54% and 100% of participants on the questionnaire scales, and were associated with performance on executive (set-shifting, working memory and planning) and non-executive short-term memory (STM) tasks. Thus, EF and STM deficits in WS could underlie behavioural difficulties in this population.

Thus, improvements in self-help skills are seen with age in both populations, but a peak in DS has been identified at age 30 followed by some decline, while in WS there is no evidence for this. Data addressing whether individuals are allowed to travel independently, or allowed out alone, indicate similar patterns, with around half of each sample not allowed to travel by themselves. Independence measures from the interview data in the DS group remained constant with age, but they increased in the WS group. There is also evidence for adaptive behaviour in general being associated with IQ in each of the groups.

It is known, therefore, that everyday functioning in WS and DS is poor, and also variable across individuals, reinforcing the heterogeneous nature of the populations (e.g., Tsao & Kindelberger, 2009). Further investigation of everyday functioning and of which cognitive skills might underlie everyday problem solving in these groups is clearly warranted. The next section addresses some of the ways in which everyday skills can be measured.

4.1.3 Measuring everyday life skills

Assessments of everyday life skills manifest at several different levels, ranging from measuring abilities experimentally (attention, EF, memory) and assuming that these processes will apply in the same way in real life, to rating scales and interviews, as described in the section above (e.g., the VABS and SDQ), which ask directly about behaviours in a real-world context. Clearly, both approaches have different merits but may not serve to inform each other, bringing us back to the ecological validity problem: experimental work does not necessarily tell us about everyday life, while descriptions of behaviour, although often clustered around certain concepts (e.g., hyperactivity/inattention and prosocial behaviour measures from the SDQ), may not tell us about underlying psychological constructs *per se*, such as inhibition or shifting (e.g., Burgess et al., 2006).

In the neuropsychological literature, there are two types of research that address real life as a response to the ecological validity problem (Spooner & Pachana, 2006). The first is research in which tasks are developed that attempt to imitate real-life situations (e.g., the Rivermead Behavioural Memory Test; B. Wilson, Cockburn, & Baddeley, 1985) including a cooking task (Frisch, Förstl, Legler, Schöpe, & Goebel, 2012). The second is research in which *bona fide* real-life situations are investigated, and the relationship with outcomes on laboratory tasks assessed (Burgess et al., 1998; Spooner & Pachana, 2006). The former is known as establishing verisimilitude and the latter, as veridicality (Franzen and Wilhelm, 1996, cited in Spooner & Pachana, 2006).

Veridicality, rather than verisimilitude, is the approach taken in the current study, using a questionnaire approach. Parental questionnaires provide an efficient way of gaining an understanding of behaviours in a real-life context. As well as allowing us to ask directly about the everyday lives of the populations of interest to assess the ecological validity of EF tasks, this approach allows us to pursue an understanding of individuals' experiences with the ill-defined problems that occur in everyday life; problem solving is examined in its natural environment.

4.1.4 Behavioral Rating Inventory of Executive Function (BRIEF)

One measure that was designed for the assessment of everyday executive functioning is the Behavioural Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000a). It takes the form of a questionnaire, asking parents directly about their child's EF. It is argued to be an ecologically valid measure with strong veridicality because it measures everyday functioning (Gioia, Isquith, Kenworthy, & Barton, 2002). The BRIEF was selected for use in the current study for several reasons. First, it provides scores on the widely recognised components of EF (including Inhibition, Shifting and Working Memory) which were measured in the study in Chapter 2 (for a list of scales see Section 4.2.4.1). This would allow comparison of experimental and questionnaire scores in this sample of participants to assess the conceptual distance between the two sets of data. Second, it has been validated and standardised using a large sample of TD children of a wide age range (5 to 18 years) indicating broad suitability for young children right through to young adults. This was important to ensure suitability, as far as possible, for the wide range of ages and abilities to be tested: parents of young TD children as well as young adults with intellectual disabilities would be completing the questionnaire. The 5 to 18 age range could also be expected to capture the mental ages of the atypical samples of adolescents and young adults. Third, the BRIEF was designed for assessment of EFs in individuals with a variety of medical conditions and disorders. Gioia, Isquith, Kenworthy, et al. (2002), the authors of the BRIEF, report data indicating different EF profiles on the BRIEF in children with ADHD subtypes, autistic spectrum disorders (ASD) and reading disorders as well as traumatic brain injury (TBI). Indeed, the professional manual also contains unpublished data for individuals with learning disabilities, making it a very relevant choice for groups with WS and DS.

Few studies have used the BRIEF in DS and WS. Lee et al. (2011) present data from 26 parents of children with DS aged between 4 and 10 years using the Preschool version of the BRIEF. Standardised scores were calculated using outcomes from IQ tests (approximating mental age, MA) instead of CA. Elevated, i.e., impaired, scores were found for Working Memory and Plan/Organize scales relative to norms for TD children with CAs that corresponded to the DS sample's MAs. In contrast, indices measuring inhibitory self-control and flexibility were not significantly elevated above estimates of the level expected for the sample's MA. As the authors acknowledge, this study does not include a control comparison group (either typical or atypical). A lack of a comparison group with another disorder or intellectual impairment will mean that conclusions cannot be drawn about the syndrome specificity of the patterns in performance that are observed.

Two studies have employed the BRIEF with a WS group. One used the measure to investigate sensory abnormalities in WS, and the other used the BRIEF to investigate the relationship between EF and anxiety diagnoses (Woodruff-Borden, Kistler, Henderson, Crawford, & Mervis, 2010). John and Mervis (2010) asked parents of children aged 6 to 10 years with WS to complete several questionnaire-based measures, including the BRIEF. On the basis of outcomes from the Short Sensory Profile (SSP), children were split into two groups: low and high sensory impairment. The 'high impairment' group scored more poorly than the 'low impairment' group on six of the eight scales of the BRIEF: Shift, Emotional Control, Initiate, Working memory, Plan/Organize, and Monitor. Unfortunately, information about the EF profile for the group that would be revealed by comparing scores across the scales was not provided. Woodruff-Borden et al. (2010) found that within a group of 33 children with WS, children with a diagnosis of an anxiety disorder scored more poorly on the Behavioral Regulation Index (BRI) than those who did not, with this difference not apparent for the Metacognition Index (MI).

4.1.5 Problem-Solving Questionnaire (PSQ)

To complement this commercially available standardised measure examining everyday EF, a novel questionnaire was designed with the specific aim of investigating everyday problem solving. In everyday life there are many opportunities for failing to solve a problem, and the aim of developing this questionnaire was to understand them in WS and DS.

The aim in mind when designing the questionnaire was to identify areas of strength and weakness in problem solving in different developmentally disordered groups, in order to further knowledge of which areas might, in future, be ripe for intervention. Strengths are important to consider as well as weaknesses and it should not always be assumed that weak areas are the areas to support when designing interventions (Karmiloff-Smith & Farran, 2012b).

4.1.5.1 PSQ Development

Several sources were consulted while the questionnaire was in its development stage. Pretz et al. (2003) note that there are seven stages of problem solving, requiring solvers to:

1) Recognise or identify the problem

2) Define and represent the problem mentally

3) Develop a solution strategy

4) Organise his or her knowledge about the problem

5) Allocate mental and physical resources for solving the problem

6) Monitor his or her progress towards the goal

7) Evaluate the solution for accuracy (p. 3).

These aspects of problem solving were used as an initial basis for the construction of the questions that parents were asked in the PSQ. Behavioural manifestations of these aspects were loosely based on these points, for example, points 5 and 6 might manifest as 'doing steps in the right order' and 'staying focused on the task'. Efforts were made throughout questionnaire construction to describe tangible behaviours that parents can make judgements about, and relevant examples were given in each different question to this end, so for getting dressed the example given for 'doing steps in the right order' is 'e.g. putting on socks before shoes' while for brushing teeth it is 'e.g. putting toothpaste on the brush before starting to brush'. In this way, attempts were made to identify any difficulties with problem solving through the behaviours that might be produced if that particular aspect of problem solving were to fail. It is also possible to speculate that specific items on the PSQ might relate to some aspects of experimental performance more than others: for example, difficulties with 'staying focused on the task' might indicate impairments in inhibition or working memory, while 'doing steps in the right order' might conceivably draw on working memory and planning abilities. However, it should be kept in mind that high-level tasks are being reported on in the questionnaires, and given the interrelated nature of EFs (e.g., Lehto et al., 2003; Miyake et al., 2000) it is likely that several aspects of EF are involved while performing real-life tasks.

Efforts were made to select everyday tasks that would be relevant for as many participants as possible. When considering everyday life, it soon became apparent that many tasks are completed as a matter of course in our daily lives, such as getting dressed and brushing our teeth. This kind of scenario had the advantage of being potentially applicable to all of the participants' lives, but being such a routine task, it seemed likely that the different steps in the task would be over learned, or automatic. By contrast, problem-solving skills really come into play when a task is novel (Rabitt, 1997b, cited in Miyake et al., 2000; Shallice, 1982). Therefore, the questionnaire was divided into two sections: four questions asked parents about their child's abilities and behaviour on everyday, or routine, tasks, and three asked about tasks that were considered more novel.

Generating novel tasks that would be encountered often enough in everyday life to allow us to ask large groups of parents reliably about the scenario was a challenge. To resolve this, tasks were chosen that, even if they were frequently encountered, would nonetheless incorporate an element of unpredictability. The novel tasks included were: finding a lost possession; packing a bag; and putting items away in a wardrobe. When finding a lost possession, the demands of the task would vary greatly depending on the item lost, location it was lost in, and so on. Packing a bag for the day would have varying demands in relation to, for example, the weather and activities ahead on that particular day. Finally, putting items away in a wardrobe/chest of drawers would vary its demands depending on how much needed to be put away and how much was already in the cupboard to be filled. In contrast, while elements of the routine tasks might also vary it was felt that these were more minor variations.

Advice from Beck (personal communication, October 13th, 2011) emphasised the importance of how children cope when something goes wrong with the problem they are attempting to solve: for example, if they are getting dressed and can't get an item of clothing on, what do they do? This advice was incorporated into the questionnaire design such that each question had an additional part: once something has gone wrong, which behaviours are children likely to exhibit? This gives a valuable window into the kinds of strategies that children use in everyday life. It also allows us to investigate how children respond when even a well-learned everyday task goes wrong, thus introducing a real element of problem solving into even the questions about routine tasks.

Presenting questionnaire development ideas to a variety of audiences also greatly aided with questionnaire design: a practitioner who interacted with parents of individuals with DS through his work noted that parents often report difficulties with brushing teeth in this group; parents at a convention for the Williams Syndrome Foundation (WSF) gave extremely useful feedback, including noting particular difficulties with keeping bedrooms tidy (prompting the inclusion of the final question regarding putting items away in a wardrobe/chest of drawers), and a colleague who is also a parent of an individual with WS reviewed the questionnaire to ascertain its appropriateness for the group and its capture of relevant aspects of problem solving. Finally, further discussion with colleagues and consultation of questionnaire design resources ("Guide to the Design of Questionnaires", n.d.; "Questionnaire Design", n.d.) and literature (Vaus, 2002) contributed to the final design of the questionnaire. A five-point Likert scale was employed, as in, for example, the “Problem-Solving Style Questionnaire” ("Problem-Solving Style Questionnaire (PSSQ)," n.d.) in order to obtain data that were quantifiable. Parents were given a choice of five responses for each item, in order to give respondents a fine-grained level of answering without overwhelming detail, and an odd number of items was chosen to allow respondents to choose the middle option if they wished, which does not indicate a direction of response.

4.2 Method

4.2.1 Participants

4.2.1.1 Demographics

Data were collected from the parents of 113 individuals in total. One participant from the DS group was excluded because the parent noted that the child also had a diagnosis of autism, leaving 112 in the data set. There was no risk of dementia in the DS group as they were aged under 30 years (cf., Lott & Head, 2001). It should be acknowledged that it was the parents who completed the questionnaires about their child’s everyday problem solving, and the parents who are therefore the respondents in the study. However, as we will be referring to the (sometimes adult) son’s or daughter’s ‘performance’ on the questionnaires, they will be hereafter referred to as the participants. Demographic information for the participants is presented in Table 4.1.

Group (N)	Males: Females	CA Range	Mean CA (SD)
		years:months	years:months (months)
TD (34)	18:16	4:10 - 11:5	8:3 (23.87)
DS (31)	14:17	10:4 - 23:9	17:4 (47.91)
WS (47)	24:23	10:7 - 26:6	18:0 (52.03)

Table 4.1: Questionnaire participant demographics

ANOVA revealed that CA was significantly different across the groups ($F(2,109) = 79.047$, $p <. 001$, partial $\eta^2 = .592$), with the TD group being younger than the atypical groups ($p < .001$ for both) and no significant difference between the ages of the WS and DS groups ($p = .787$). Data regarding the presence or otherwise of siblings were available for 16 participants in the DS group and 30 in the WS group. Of these, two individuals with DS and four with WS were known to not have any siblings.

4.2.1.2 Mental age measures

Additional data regarding participants' verbal and nonverbal ability were also available for a subset of each group because they had participated in the study in Chapter 2. These data are summarised in Table 4.2 below.

Group (N)	Males: Females	CA Range	CA Mean (SD)	RCPM Raw Score	BPVS Raw Score
		years: months	years: months (months)		
TD (25)	11:14	4:10 - 11:5	8:3 (24.36)	26.40 (6.92)	108.28 (22.95)
DS (14)	8:6	12:2 - 23:9	18:1 (45.12)	18.29 (6.09)	82.50 (28.06)
WS (16)	8:8	12:2 - 22:10	17:6 (39.72)	17.69 (3.32)	115.19 (24.13)

Table 4.2: Demographics of subsets of questionnaire participants with mental age data[4]

For these subsets, ANOVAs revealed that performance on BPVS and RCPM differed across the three groups. On the BPVS, there was a main effect of Group ($F(2,52) = 7.346$, $p = .002$, partial $\eta^2 = .220$). The DS group performed more poorly than the WS ($p = .006$) and TD ($p = .020$) groups, with no reliable difference between WS and TD ($p = .637$). For the RCPM scores, the main effect of Group ($F(2,52) = 14.061$, $p < .001$, partial $\eta^2 = .351$) was driven by reliably better performance in TD than DS ($p = .002$) and in TD than WS groups ($p < .001$) with no significant difference between the two atypical groups ($p = .943$). Given the similar age ranges of the subset with mental age data to the larger set for whom questionnaire data are available, it can be tentatively assumed that the subsets are representative of the larger groups in terms of their RCPM and BPVS scores.

CA was significantly related to BPVS score in each group (TD: $r = .870$, $p < .001$; DS: $r = .622$, $p = .018$; WS: $r = .694$, $p = .003$). CA was significantly related to RCPM score in the TD ($r = .790$, $p < .001$) and DS groups ($r = .717$, $p = .004$) but not in the WS group ($r = -.198$, $p = .463$).

[4] In order to ensure that the slightly wider age range in the questionnaire group than in the experimental group was not affecting outcomes, questionnaire performance was compared in the experimental subsets and was found to be consistent with the results pattern for the larger questionnaire groups, seen in Figure 4.3 (WS < DS < TD). Thus, the pattern of questionnaire performance against the backdrop of the BPVS and RCPM scores is supported, and it is unlikely that individuals at the extremes of the questionnaire respondents' age range were unduly affecting the outcomes.

4.2.1.3 Recruitment

In the case of the TD group, questionnaires were sent out to parents of the children who participated in the experimental study via the school office, plus some additional parents of children from each age group. Parents of the atypical participants were either posted questionnaires to complete and return in a Freepost envelope or, in the case of some of the atypical participants, parents completed them while their child was the experimental tasks (Chapter 2). On occasion one or both questionnaires were completed by telephone, although written consent was always obtained. Parents of atypical participants were either parents who had previously worked with the research group or were recruited anew through the Williams Syndrome Foundation (WSF) or Down Syndrome Association (DSA). In the case of WS, parents were only recruited if their child had a diagnosis of WS confirmed both phenotypically and by a genetic fluorescence *in situ* hybridisation (FISH) test for the elastin deletion. For both atypical groups, efforts were made to ensure that questionnaires were sent to parents in a range of geographical areas, with some parents from the London area and some outside the London area, in order to sample a range of socio-economic status (SES) families. Informed consent was obtained from all respondents. The overwhelming majority of respondents were mothers (96) with 10 responses from fathers (hence, the respondent group is referred to as 'parents') and a small number of other respondents, e.g., stepmother or sister (supervised by parents). One mother completed the questionnaire together with support workers where her son lived at a nearby care home. Most of the sets of questionnaires were completed by the same person, but on exception the respondents were different, e.g. for one participant their mother completed one questionnaire and their father, the other.

4.2.2 Materials

Parents completed two questionnaires: the Behavioral Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000a) and our novel Problem-Solving Questionnaire (PSQ).

4.2.3 Procedure

Parents received an envelope containing both questionnaires, an information sheet about the study, and an additional information sheet concerning the BRIEF, asking parents to omit any items that they felt were not relevant for their child. The BRIEF consists of a list of 86 behavioural descriptions, or items, and for each item, parents are asked to select one of three options: whether the behaviour has Never, Sometimes or Often been a problem for the child over the last six months. On the website ("BRIEF", "BRIEF (Behavior Rating Inventory of Executive Function),") the form is reported as taking 10-15 minutes to complete.

The PSQ consists of seven questions asking about different problems that their child might encounter in everyday life. For the majority of the questionnaire parents respond by choosing between several options and circling the relevant one. Each question also includes the opportunity for a more open-ended response. At the end of the everyday tasks and before the novel problems section there is space for parents to describe any strategies that have been helpful for their child when performing everyday tasks, and to report whether this strategy was taught or self-initiated. There is also space at the end for any further comments. The questionnaire is estimated to take approximately 30 minutes to complete, depending on the level of detail provided by the parent.

4.2.4 Design

4.2.4.1 The BRIEF

The BRIEF provides a score on eight scales of executive functioning: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials and Monitor. Combining the first three scales gives a score on a Behavioural Regulation Index (BRI) and the latter five, on a Metacognition Index (MI). These two indices further combine to give a global score: the Global Executive Composite (GEC). The BRIEF is a standardised measure with acceptable reliability and validity (for more information, see Gioia, Isquith, Guy, & Kenworthy, 2000b).

4.2.4.2 The PSQ

The PSQ was developed for the purpose of this study in order to understand problem solving in everyday life. It takes a different approach to the BRIEF in that, rather than asking about how often behaviours occur, parents are asked to rate how easy their child finds different aspects of several tasks, and then asked to imagine that something has gone wrong with the task and asked how likely their child might be to produce certain responses. There were also several opportunities for parents to provide more open-ended responses.

4.2.4.2.1 PSQ design

The PSQ can be found in Appendix D. It consists of seven questions: the first four ask about routine tasks and the remaining three, about more novel problems. A list of question topics can be found in Table 4.3.

Question Number	Routine/Novel	Topic
1	Routine	Getting dressed
2	Routine	Brushing teeth
3	Routine	Making a sandwich
4	Routine	Making a telephone call
5	Novel	Finding a lost possession
6	Novel	Packing a bag for the day
7	Novel	Putting items away in a wardrobe/chest of drawers

Table 4.3: PSQ question topics

Each of the seven questions follows the same format, with some variations between questions, as appropriate. Each question is split into two parts, ‘approach’ and ‘response’, with various aspects of problem solving addressed in each part: see Table 4.4 for a summary of aspects that were addressed in each question, along with an abbreviated aspect name that will be used throughout the chapter.

	Aspect	**Abbreviation**
Approach		
	Knowing what the goal of the problem is	Know
	Routine: Knowing what can be done to reach the goal Novel: having some ideas about what could be done to reach the goal	What
	Routine: Doing steps in the right order Novel: Keeping track of what they are doing	Steps
	Staying focused on the task	Focus
	Stopping when a task is finished	Stop
Response		
	Change what they have been doing in response to the problem	Change
	Ask for help	Ask
	Become Emotional	Emotion
	Lose focus	Lose focus
	Stop their efforts through lack of perseverance	Lack perseverance

Table 4.4: Aspects of problem solving shared across questions in the PSQ

While the majority of the items appeared across all the questions and are presented here, certain items only appeared in some questions, for example, when asking about finding a lost possession, an item was included to assess how likely the child is to ‘end their search after a sensible amount of time if the item has not been found’. For full details of the contents of each question, see Appendix D.

At the beginning of each question about routine tasks, parents noted whether their child usually does the task independently; with some help or only if someone else does it for them; if 'some help' was chosen they are asked whether the help was in the form of specific instructions, general encouragement or physical support. For the Getting Dressed question they also noted whether the child chooses the clothes themselves or whether someone else chooses them, as this was felt to be a salient potential reason for failing the task. While getting dressed and brushing teeth are extremely likely to happen daily, making a sandwich and making a telephone call are conceivably less so; therefore at the start of these questions parents rated how often their son or daughter attempts the task.

Then, the approach section of each question asks about how easy a child would find each aspect of a task, and parents are asked to circle a number between 1 (very easy) and 5 (very difficult). This constitutes a measure of the child's capability with the task. In the response part of the problem, parents are asked how likely their child would be to display certain behaviours if something had gone wrong with the task, for example when getting dressed, a t-shirt was put on backwards. Parents choose between five options again, with 1 representing very likely and 5, very unlikely. The next part of each question addresses any other behaviour the child might be likely to exhibit, thus giving the parent the opportunity to provide any additional information they feel is necessary. Finally the parent was asked what the reason was likely to be, if their son or daughter were to fail the task.

4.3 Results

Results of the PSQ and BRIEF questionnaires will be presented through: the comparison of performance across groups on an overall percentage score for each questionnaire; an analysis of chronological and mental age measures and how they relate to overall performance in each group; and subsequently, a detailed analysis of performance on each questionnaire. First, the data handling and scoring considerations for the two questionnaires (which are used in Sections 4.3.3 and 4.3.4) are outlined, in order to illustrate the way in which the overall percentage scores were calculated. For the open-ended parts of the PSQ questions, a large number of response areas were left blank. This has not been analysed.

4.3.1 Data handling

4.3.1.1 BRIEF

4.3.1.1.1 Scoring

The BRIEF consists of 86 statements for which parents are asked to judge whether the behaviour is Never, Sometimes or Often a problem, scoring one, two and three points respectively. Different numbers of items contribute to each of the eight scales (in Section 4.2.4). The sum of the points for all the items in a scale gives a raw score for the scale. There are also 14 additional items which are classified as falling under one of the scales but do not contribute to the raw score for any scale. Raw scores can be converted into T scores with a mean of 50 and a standard deviation of 10 (see Section 4.3.3.3). One approach to scoring on the BRIEF, used particularly when dealing with group data (e.g., as used by Gioia et al., 2000a) is to calculate the mean item score (between 1 and 3) for a scale. This allows comparison of scales which vary in the number of items that they contain (Gioia et al., 2000a p.66). This is therefore the approach adopted in the BRIEF analysis section (Section 4.3.3).

4.3.1.1.2 Missing data

The BRIEF is designed for administration to children aged between 5 and 18 years, and includes several questions pertaining to school and homework. As many of the participants in the study were older than 18 years and as even the younger participants may not be engaging with, for example, homework, in the same way as a typical population (e.g., planning for long-term assignments) instructions were given to parents on administration to not answer any question that they felt was not relevant for their child.

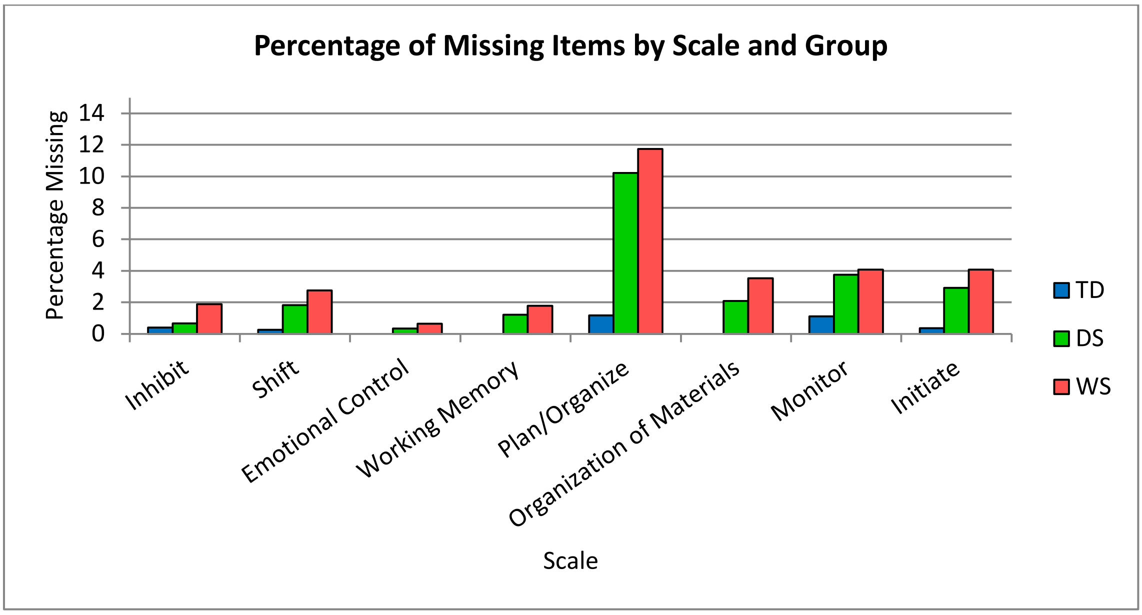

Figure 4.1: Percentage of missing items by BRIEF scale and group

Figure 4.1 summarises, for each group, the number of missing items for each scale as a percentage of the total number of potential responses for that scale, including the additional items that do not contribute towards raw scale scores.

A striking pattern when viewing Figure 4.1 is the high number of missing items in the DS and WS groups for the Plan/Organize scale. Out of 14 items, ten were related to school life. Only four remained that seem to apply outside school. For the reasons discussed in the paragraph above, the link to school life is likely to account for the high levels of missing responses for the atypical groups.

To check for syndrome-specific patterns of missing data, the percentage of respondents omitting each item was calculated separately for each group. Those items with a percentage omitted falling two or more standard deviations above the mean were identified. The vast majority of items omitted were in the Plan/Organize scale (corresponding to Figure 4.1). Five items met this criterion for the WS group and six for the DS group, with an overlap between groups of three items. All were related to school life and thus do not reflect syndrome-specific patterns in response.

In order to deal with missing data (which were a natural consequence of the instructions given to parents to omit irrelevant items) the average of all of the responses provided, across all scale items (not including the 14 additional items) and across all participants, was used to replace all the missing data points. This was judged to be the most conservative approach to dealing with missing data, as it minimises the differences in scores across groups, so that any group differences that are found would be likely to be robust.

4.3.1.1.3 Consistency

Using the BRIEF manual it is possible to calculate an inconsistency score for each respondent. This is achieved by calculating the absolute difference in scores between ten particular pairs of items with high correlations between the pairs of scores (with r values between .64 and .84) in their clinical sample of 852 children. The sum of the ten absolute differences gives the inconsistency score, which ranges from 0 to 20. Scores less than or equal to 6 are deemed acceptable, 7 to 8 questionable, and 9 or more classified as inconsistent. Of the current sample, one parent in the TD group scored 7, one DS parent scored 8, and three WS parents scored 7, while all of the remaining scores fell in the 'acceptable' category at 6 or below. The average score for each group was calculated and compared using a one-way ANOVA, which did not reveal any significant differences across the groups ($F<1$).

4.3.1.2 PSQ

4.3.1.2.1 Scoring

Parents responded to the PSQ by selecting one of five Likert-style options, ranging from an answer of 1 representing 'Very easy' or 'Very likely' while 5 represented 'Very difficult' or 'Very unlikely'. In the main, answers of 1 scored four points; answers of 2, three points; 3, two points; 4, one point; and 5, zero points. Thus, being more likely to achieve elements of problem solving was associated with better scores, e.g., when asked how likely their son or daughter would be to change what they were doing in response to a problem, a response reporting their being very likely to do this would achieve the most points (four), with 'Very unlikely' scoring the fewest points (zero). However, in the case of three of the variables with a negative valence ('become emotional', 'lose focus' and 'stop their efforts due to a lack of perseverance') the scores were reversed such that the most points were awarded when participants were reported as the least likely to display these behaviours. This served to preserve the direction of scoring, so that higher scores were always equated with better problem solving.

4.3.1.2.2 Missing data

Responses to the PSQ did not follow a uniform pattern in terms of the response rate per question. Overall, questions 1, 2, 5, 6 and 7 were answered by more than 98% of parents. Questions 3 and 4, however, saw a much lower response rate, of 73.21% and 82.14% respectively. This is illustrated in Figure 4.2, displaying, for each group, the percentage of respondents for each group omitting the whole question (reminders of the question topics are included below Figure 4.2). The rightmost cluster of bars displays the number of items omitted within whole questions that were completed, expressed as a percentage, for the group, of the total maximum number of responses. That is:

$$\frac{\text{Total number of items omitted, not constituting omitted whole questions}}{\text{N (number of respondents in group)* 95 (number of items on PSQ)}}.$$

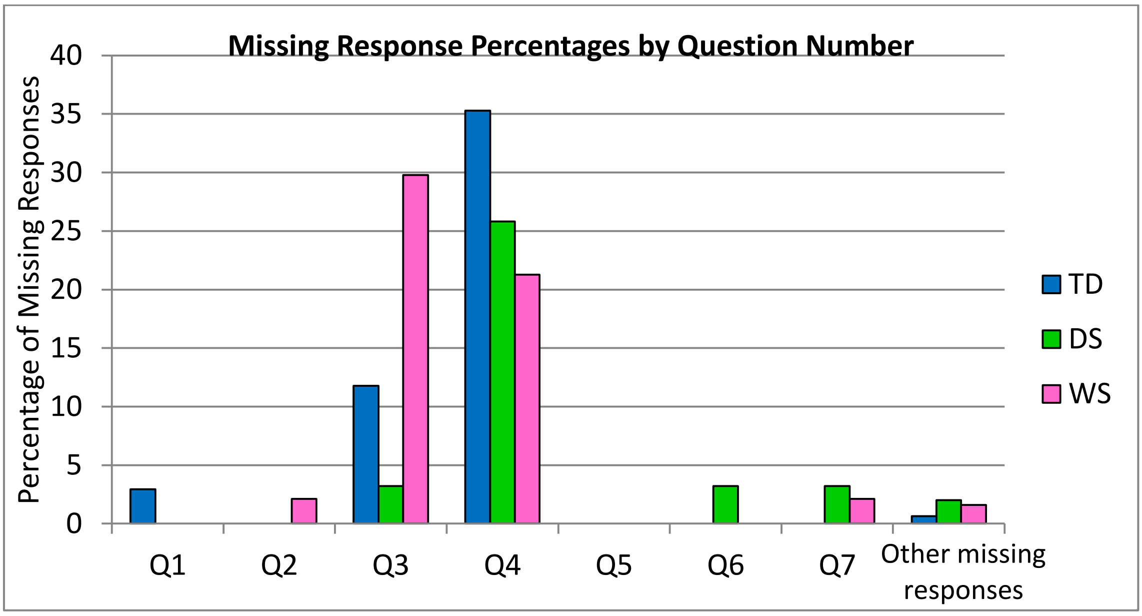

Figure 4.2: Percentage of missing responses by question on the PSQ

Q1	Getting dressed	Q5	Finding a lost possession
Q2	Brushing teeth	Q6	Packing a bag for the day
Q3	Making a sandwich	Q7	Putting items away in a wardrobe/chest of drawers
Q4	Making a telephone call		

Given the large numbers of missing responses to questions 3 and 4, overall analysis of the PSQ will include questions 1, 2, 5, 6 and 7 only. As for the BRIEF analysis, all missing data points were replaced with the overall mean of all the responses, across all groups.

4.3.1.3 Approaches

4.3.1.3.1 Selection of an appropriate *post hoc* comparison procedure

Given the unequal sample sizes across groups, for between-participants *post hoc* tests, the Games-Howell procedure was selected over Tukey because it is accurate for unequal sample sizes (as well as taking account of unequal population variances) (Field, 2009). Thus it is reported in the analysis above and throughout the chapter. Where data did not meet the assumption of sphericity, the adjusted F value and degrees of freedom are reported according to the Greenhouse-Geisser correction.

4.3.1.3.2 Likert-type scales

Both questionnaires use a Likert-type scale. Grace-Martin (2008) notes that researchers disagree on the use of Likert-style data in parametric tests; some argue that it is not appropriate because of the ordinal nature of the data. She makes several recommendations: a minimum of 5 points used in the scales, caution with results that are close to the significance boundary and the use of non-parametric equivalent tests where possible. Given these considerations, nonparametric tests were used in parallel to parametric tests, and are reported alongside them when they show a different pattern of results.

4.3.2 Percentage score analysis

4.3.2.1 Calculation

In order to give an overall indication of reported performance on each questionnaire, a percentage score was calculated for the PSQ and the BRIEF separately. Given that respondents were instructed to leave out any items that they felt were less relevant, it was important not to introduce penalties for missing items in the scoring. Therefore, scores were calculated using the following formula, for data from all questions (1-7):

$$\frac{\text{Points awarded}}{\text{Maximum possible points based on the number of responses made}} \times 100\%.$$

Thus, individual respondents were not penalised for making fewer responses.

4.3.2.2 Assigning points

The points awarded for each response for the calculation of percentage scores were decided upon in order to achieve consistency between the two questionnaires. The scoring on the PSQ, as detailed in Section 4.3.1.2.1, is arranged so that the most positive response earned four points and the most negative, zero, with three intermediate options earning three, two and one points. When completing the BRIEF, parents selected one of three options for each statement: whether the behaviour was Never, Sometimes or Often a problem for their child. In the same way as for the PSQ, the best response possible (Never) was assigned four points for the percentage score calculation; Sometimes, two points; and Often, zero points. Very rarely, parents would circle two adjacent answers, e.g., Often and Sometimes, and in these instances the score given reflected this, being midway between the score for each circled answer: in this case, one point (midway between zero points for Often and two points for Sometimes). Higher percentage scores thus equated with better problem solving, for both questionnaires.

4.3.2.3 Group comparisons

Groups were compared on the overall percentage scores for each questionnaire, as illustrated in Figure 4.3.

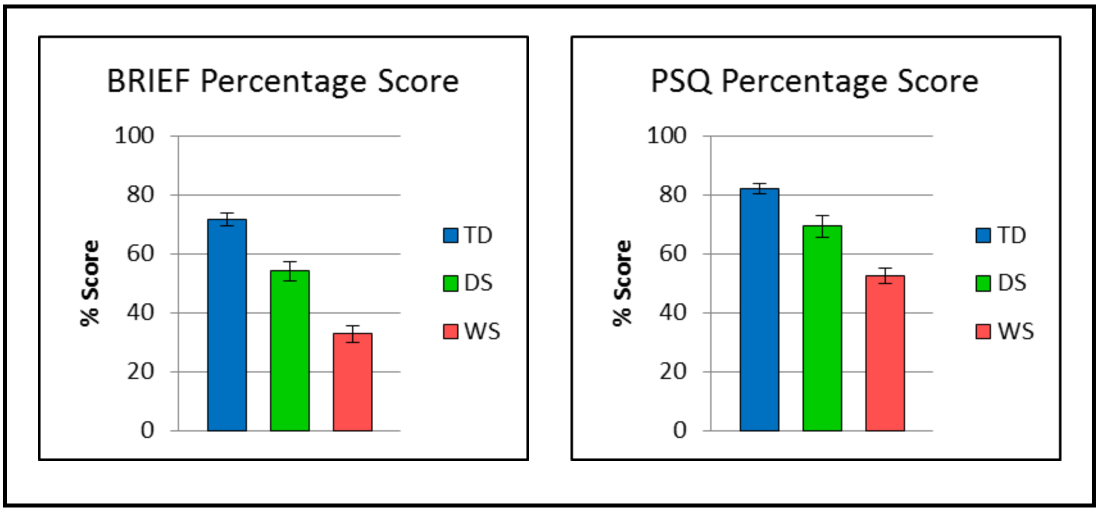

Figure 4.3: Overall percentage scores on the BRIEF and PSQ by group

As a reminder, DS and WS groups were approximately matched on nonverbal IQ (Raven's score) while the TD group had a higher nonverbal IQ but was significantly younger. A mixed ANOVA on the overall percentage scores on the two questionnaires by Group revealed a significant main effect of Questionnaire ($F(1,107) = 148.793$, $p < .001$, partial $\eta^2 = .582$), with percentage scores being higher for the PSQ than the BRIEF overall. A significant main effect of Group ($F(2,107) = 47.065$, $p < .001$, partial $\eta^2 = .468$) indicated higher scores for TD than DS groups ($p = .001$) and for DS than WS ($p < .001$) using the Games-Howell procedure. There was a significant interaction between Questionnaire and Group

(p = .006, partial η^2 = .090). One-way ANOVAs revealed main effects of Group for each questionnaire (p < .001 for both) with Games-Howell comparisons indicating better performance for TD than DS groups (PSQ: p = .016; BRIEF: p < .001), and for DS than WS groups (PSQ: p = .002; BRIEF: p < .001). Equally, paired samples t-tests indicate significant within-group differences for each questionnaire (p < .001 for all). Thus, the interaction appears to be due only to degrees of difference; there are greater group differences on the BRIEF than on the PSQ. For each group, the two overall percentage scores were significantly correlated with each other (TD: r = .552, p = .001; DS: r = .806, p < .001; WS: r = .703, p < .001).

Small numbers of the WS and DS group were known not to have any siblings. To assess whether the presence of siblings influenced parental questionnaire ratings, the percentage score on each questionnaire was compared between children with and without siblings. Because the number of individuals known to be without any siblings was so small (two in DS and four in WS) nonparametric tests were used in this analysis. For both groups, Mann-Whitney tests indicated no significant differences in percentage scores between individuals with and without siblings (p > .05 for all). In addition, all the scores for individuals without siblings fell within two standard deviations of the mean of the percentage scores of the individuals with siblings.

4.3.2.4 Developmental trajectories

4.3.2.4.1 CA

As the trajectories for the typical and atypical groups are non-overlapping, the CA trajectories were compared for the WS and DS groups only. CA and percentage scores on the BRIEF are presented in Figure 4.4.

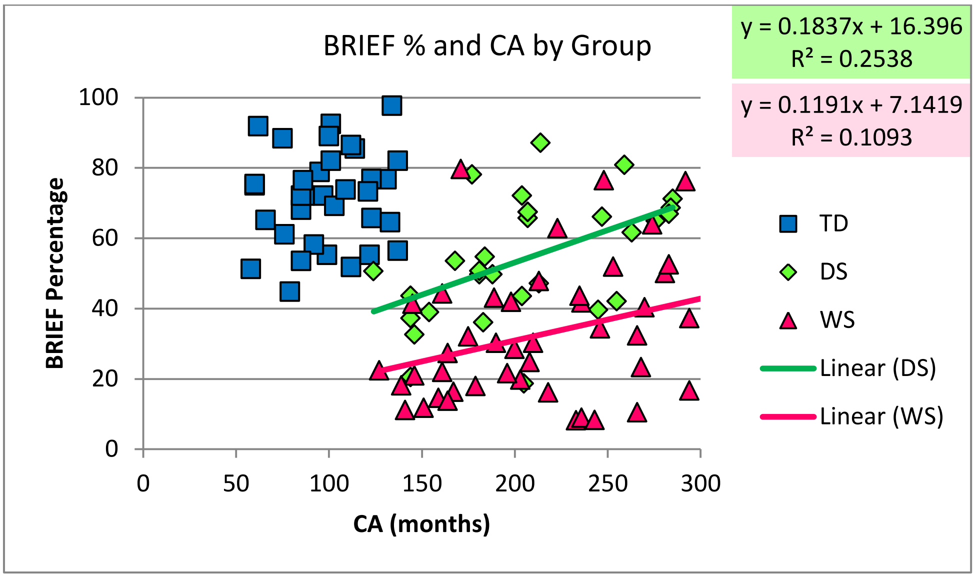

Figure 4.4: Developmental trajectories of BRIEF percentage based on CA

CA was significantly related to BRIEF percentage score in each of these groups (but not in the TD group) (DS: r = .504, p = .005; WS: r = .331, p = .025; TD: r = .116, p = .515). ANCOVA was conducted on the WS and DS data, with CA as the predictor variable, with the CA by BRIEF score interaction term included in the model. ANCOVA revealed a significant difference between the groups at the lowest level of CA (p = .030) and a significant effect of the covariate, CA (p < .001). The gradient of the slopes did not differ significantly (F<1) indicating that the relationship between CA and BRIEF score did not differ between the two atypical groups. The scatter plot for the PSQ is displayed in Figure 4.5.

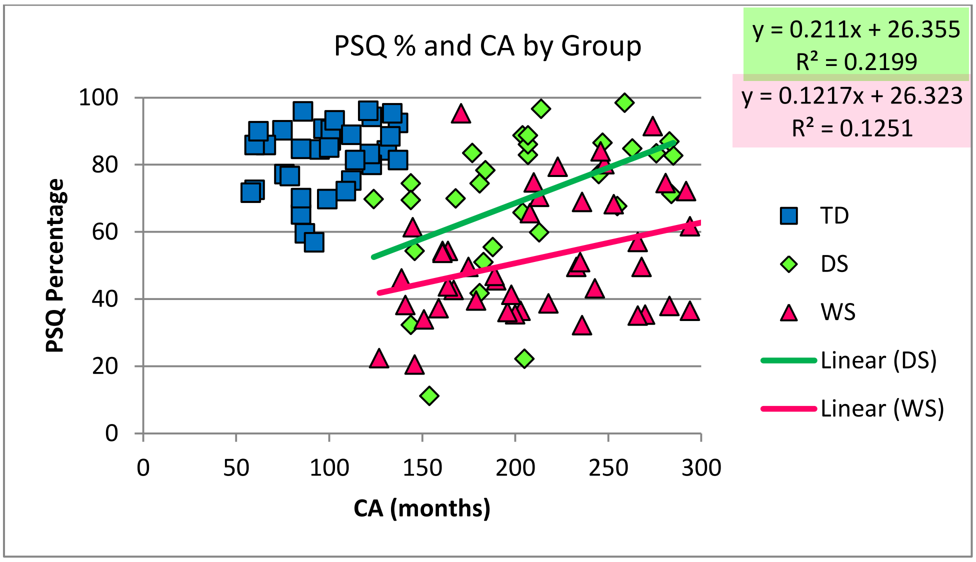

Figure 4.5: Developmental trajectories of PSQ percentage based on CA

Correlations between CA and PSQ percentage score were significant for the atypical groups (DS: r = .426, p = .017; WS: r = .334, p = .022) and marginally related in the TD group (r = .311, p = .074). ANCOVA on the PSQ data revealed no significant differences between the two atypical groups' scores at the lowest level of CA (p = .155). While PSQ scores increased with CA overall (p = .001) the two slope gradients did not differ significantly (F<1), indicating that this relationship was similar for DS and WS groups.

Finally, for those participants for whom mental age data were available (demographics are provided in Table 4.2), the relationships between verbal and nonverbal MA and questionnaire scores were examined by comparing the developmental trajectories of the BRIEF and PSQ overall percentage scores, using BPVS and RCPM scores as predictors.

4.3.2.4.2 BPVS

Figure 4.6 displays the scatter plot of the BRIEF and BPVS data.

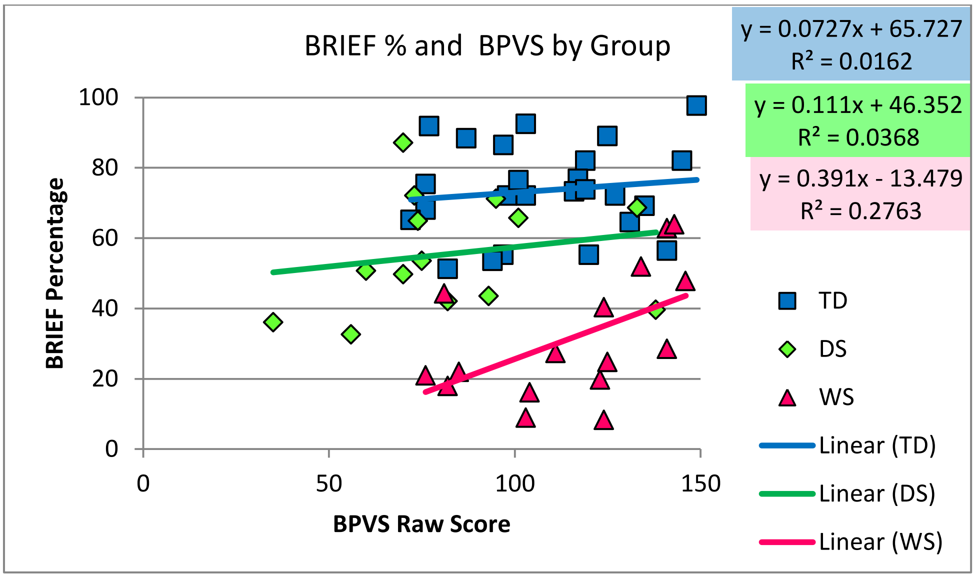

Figure 4.6: Developmental trajectories of BRIEF percentage based on BPVS score

The correlation between the two measures was significant in WS (r = .526, p = .037) but not in the other two groups (TD: r = .127, p = .545; DS: r = .192, p = .511). ANCOVA revealed that the BRIEF scores differed across groups at the lowest BPVS scores ($F(2,49) = 18.975$, $p < .001$, partial $\eta^2 = .436$), with WS performance being lower than DS performance, which was lower than TD performance (p < .05 for all). While BPVS was related to BRIEF score overall ($F(1,49) = 5.112$, $p = .028$, partial $\eta^2 = .094$) the difference in the slopes did not reach significance ($p = .279$). The scatter plot for the BPVS and PSQ data is displayed in Figure 4.7.

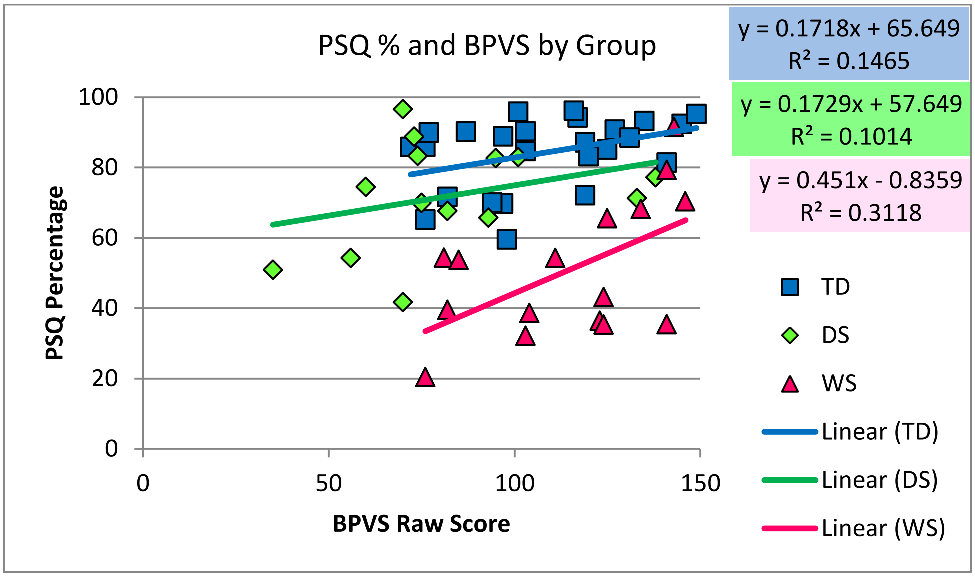

Figure 4.7: Developmental trajectories of PSQ percentage based on BPVS score

There was a significant association between BPVS and PSQ percentage overall ($F(1,49) = 12.093$, $p = .001$, partial $\eta^2 = .198$). Correlations between BPVS and PSQ scores were significant in the WS group ($r = .558$, $p = .025$), marginally significant for the TD group ($r = .383$, $p = .059$) and non-significant in the DS group ($r = .318$, $p = .267$). At the lowest levels of BPVS score, the PSQ scores differed across groups ($F(2,49) = 16.788$, $p < .001$, partial $\eta^2 = .407$), with *post hoc* ANOVAs revealing lower WS scores than each other group ($p < .001$ for both) and no difference between DS and TD groups ($p = .141$). Overall, the BPVS slopes did not vary significantly ($p = .265$). Figure 4.8 displays the scatter plot for the RCPM and BRIEF.

4.3.2.4.3 RCPM

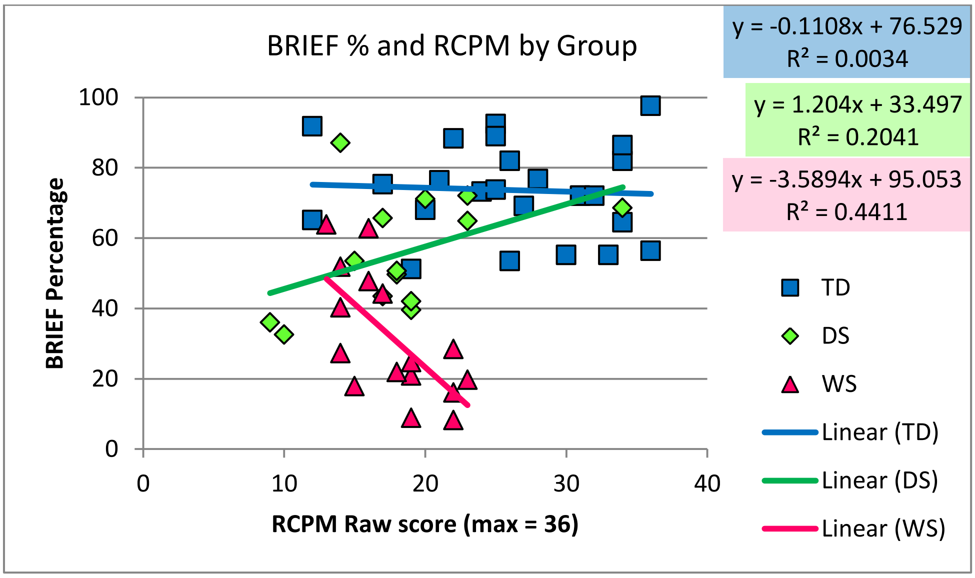

Figure 4.8: Developmental trajectories of BRIEF percentage based on RCPM score

The association between RCPM and BRIEF scores overall did not reach significance ($F(1,49) = 3.557$, $p = .065$, partial $\eta^2 = .068$) while correlations between BRIEF percentage and RCPM score in each group revealed no significant correlations for TD and DS groups (TD: $r = -.058$, $p = .781$; DS: $r = .452$, $p = .105$) but a significant negative correlation in WS ($R = -.664$, $p = .005$). ANCOVA revealed that the groups' BRIEF scores differed at the lowest level of RCPM ability ($F(1,49) = 3.557$, $p = .003$, partial $\eta^2 = .208$) and that there were differences in the rate of BRIEF outcome development with RCPM score between groups ($F(2,49) = 7.271$, $p = .002$, partial $\eta^2 = .229$). *Post hoc* ANCOVAs revealed, at the lowest levels of RCPM score, higher TD than DS and WS BRIEF scores ($p = .003$ for both) and no difference between DS and WS groups' scores ($p = .927$). The rate of change of the BRIEF score with RCPM ability was different between the WS group and the other two groups (DS: $p = .001$; TD: $p = .004$) and marginally different between TD and DS groups ($p = .092$). While the TD and DS groups' trajectories converged, with increasingly similar BRIEF scores with increasing RCPM ability, upon examining the scatter plot it is clear that the WS trajectory diverged from that of the other two groups with increasing RCPM scores. RCPM and PSQ scores are displayed in Figure 4.9.

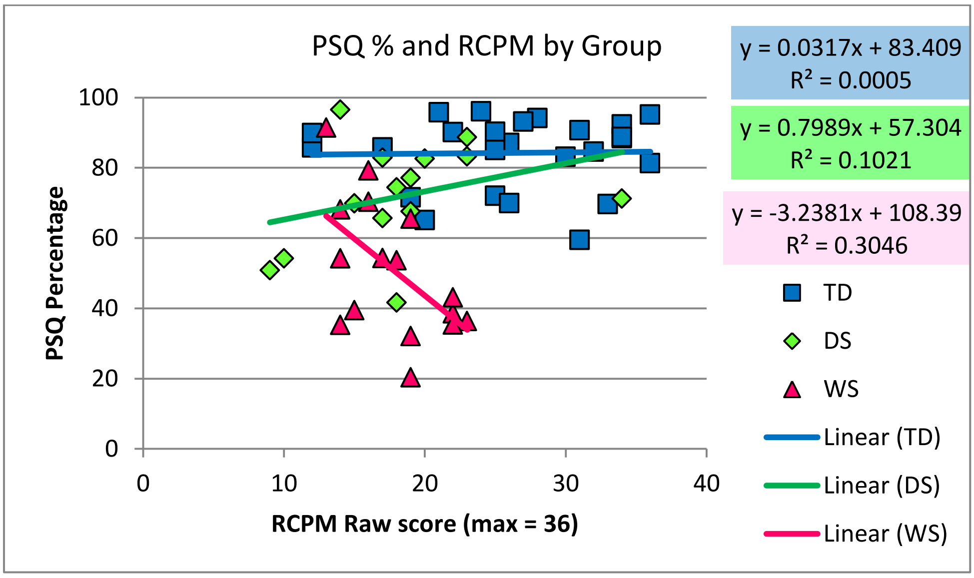

Figure 4.9: Developmental trajectories of PSQ percentage based on RCPM score

Overall, the association between RCPM score and PSQ percentage was marginally significant (F(1,49) = 3.431, p = .070, partial η^2 = .181) while correlations for each group indicated a significant relationship for the WS group (r = -.552, p = .027) but not for the other two groups (TD: r = .021, p = .919; DS: r = .320, p = .265). Overall, the PSQ scores at the lowest level of RCPM ability differed marginally across groups (F(2,49) = 2.720, p = .076, partial η^2 = .100) with *post hoc* ANCOVAs indicating higher TD than DS scores (p = .027) and higher TD than WS scores (p = .041) with no difference between DS and WS scores (p = .879). The slopes differed by group (F(2,49) = 5.416, p = .007, partial η^2 = .181). *Post hoc* analysis demonstrated that there was no difference in slope between TD and DS trajectories (p = .256), while the WS trajectory again differed significantly from that of the other two groups (DS: p = .010, TD: p = .005), diverging with increasing RCPM score.

4.3.3 BRIEF analysis

Two respondents (one from the DS group and one from the WS group) only completed the PSQ. Thus the Ns for the BRIEF data were 30 (DS), 46 (WS) and 34 (TD).

4.3.3.1 BRIEF scales

Figure 4.10 displays the mean item score per group for each scale.

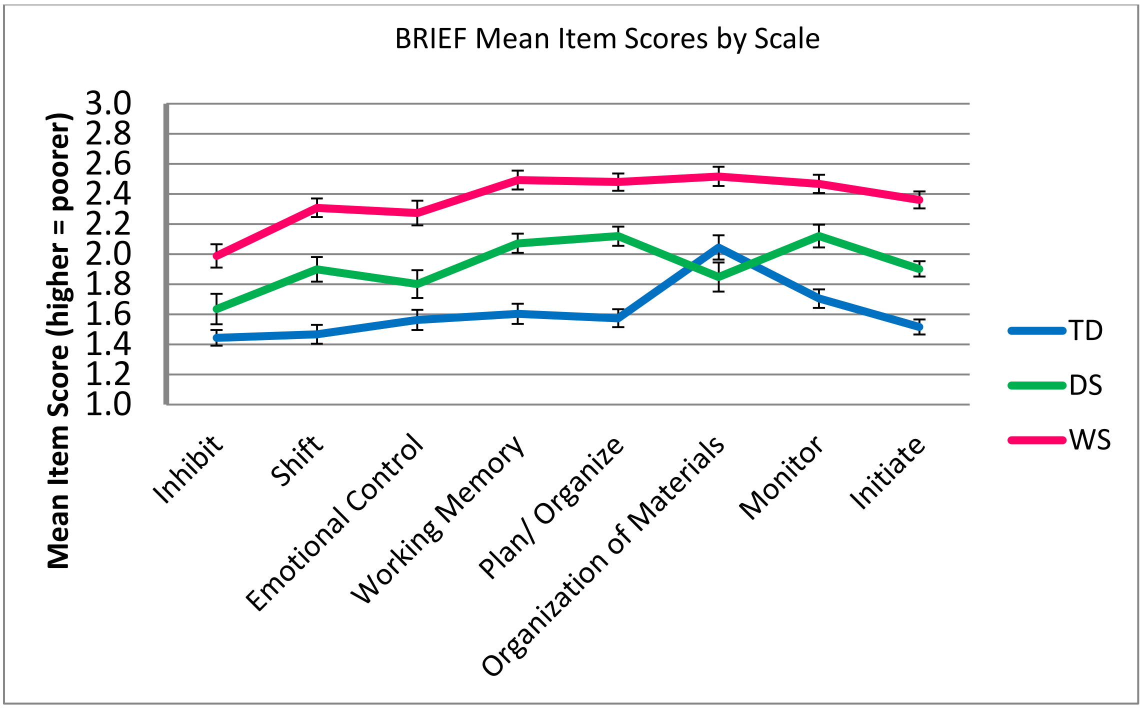

Figure 4.10: Mean item score on the BRIEF scales by group

ANOVA on the mean item scale scores revealed a significant main effect of Group ($F(2,107) = 52.141$, $p < .001$, partial $\eta^2 = .494$). *Post hoc* tests revealed better performance in the TD group than the DS group ($p = .001$) and WS group ($p < .001$), with better scores in the DS group than the WS group ($p < .001$). There was a significant main effect of Scale ($F(5.31,568.63) = 27.455$, $p < .001$, partial $\eta^2 = .204$). This is explored with reference to the significant interaction between Scale and Group ($p < .001$). One-way ANOVAs indicated that Group effects were significant for each scale ($p < .001$ for all). *Post hoc* Games-Howell tests indicated that for the Shift, Working Memory, Plan/Organize, Monitor and Initiate scales, the TD group outperformed the DS group and the DS group outperformed the WS group ($p < .05$ for all). For the remaining three scales (Inhibit, Emotional Control and Organization of Materials) TDs outperformed the WS group ($p < .001$ for all) and did not differ significantly from the DS group ($p > .05$ for all). The DS group also outperformed the WS group on these three scales ($p < .05$ for all). In other words, participants with DS scored more successfully than participants with WS on every scale. On some scales DS performance was in line with the TD group's, and on other scales it was poorer. Interestingly, Organization of Materials appeared to be noticeably poor in the TD group. Repeated measures ANOVAs of the effect of scale for each group supported this. Whilst the profile of scale scores was broadly similar across groups, the Organization of Materials scale exhibited unusual results. In TD it was poorer than each other scale. In WS it was scored less successfully than Inhibit, Shift, Initiate and Emotional Control. However, in DS it scored better than Inhibit, Working Memory, Plan/Organize and Monitor.

4.3.3.1.1 Organization of Materials: syndrome specificity

To compare the atypical group performance only, the mean item scores were recalculated, this time replacing missing scores with the average of all the responses to the scale items (i.e., the first 72) across the atypical groups. The Scale ($F(5.22, 385.88) = 22.058$, $p < .001$, partial $\eta^2 = .230$) and Group effects ($F(1,74) = 29.298$, $p < .001$, partial $\eta^2 = .284$) persisted, with better performance in DS than WS at each scale ($p < .05$ for all). The Scale by Group interaction also remained significant ($p = .048$). When removing the Organization of Materials scale, the interaction was no longer significant: $p = .706$. The interaction also remained significant with each of the other scales removed ($p < .05$ for all) with the exception of Monitor, without which the interaction term did not reach significance ($p = .060$). Thus, the Organization of Materials scale was the best marker of syndrome-specificity, but also showed an unexpected pattern for the TD group.

4.3.3.1.2 Exploring the Organization of Materials scale

Some further analyses were run in order to further the understanding of this pattern of results. First, individual items comprising the Organization of Materials scale were examined. There are six items in the scale (without additional items), of which four include the word 'mess' or derivatives of it, and all items imply that a physical space is left in an unordered way by the individual. ANOVA of Group (TD, DS, WS) by Item (6 levels) demonstrated a significant main effect of Group ($F(2,107) = 20.209$, $p < .001$, partial $\eta^2 = .274$), and of Item ($F(4.506,482.138) = 6.885$, $p < .001$, partial $\eta^2 = .060$), but no significant Group by Item interaction ($p = .186$), so the pattern of results cannot be attributed to any particular item.

Second, potential links between the Organization of Materials scale score and other background measures were ascertained to illustrate potential factors accounting for the unusual results. Correlations between CA and Organization of Materials scores were non-significant for each group (p >.05 for all), providing no evidence for development of this ability with age. In the BRIEF manual (Gioia et al., 2000a) it is noted that "The Organization of Materials scale assesses the manner in which children order or organize their world and belongings" (p. 20) which would presumably involve nonverbal reasoning. The correlation between RCPM and Organization of Materials was marginally significant in WS (r = .483, p = .058) and TD (r = .338, p = .099) groups, but not DS (p = .880) although the Spearman's correlation (run due to the use of Likert-style data) did not yield a significant correlation in the WS group (r = .419, p = .106). BPVS was not related to this scale score for any group (p > .05 for all).

4.3.3.1.3 Discriminant Functions Analysis (DFA)

DFA was conducted on the eight mean scale scores to assess whether the groups could be differentiated based on their parental BRIEF responses. The analysis identified two functions, the first of which explained 84.9% of the variance, canonical R^2 = .624, and the other, 15.1% of the variance, canonical R^2 = .228. Both functions together significantly differentiated the groups, L = .290, $X^2(16)$ = 128.224, p < .001, and after removing function 1, function 2 still significantly differentiated the groups alone, L = .772, $X^2(7)$ = 26.771, p < .001. In terms of correlation coefficients between variables and functions, Organization of Materials loaded most highly onto function 2, while the remaining seven of the variables had their highest loading on function 1. The function loadings, or correlation coefficients, are displayed in Table 4.5, in order of the size of the correlation. Coefficients in bold are the highest for each scale.

	Function 1	Function 2
Initiate	**.826**	.425
Plan/Organize	**.793**	.120
Working Memory	**.736**	.242
Shift	**.664**	.240
Monitor	**.626**	.182
Emotional Control	**.458**	.378
Inhibit	**.367**	.284
Organization of Materials	.296	**.887**

Table 4.5: Factor loadings of BRIEF scales for discriminant functions analysis

Function 1 discriminated the TD from the WS and DS groups, while function 2 discriminated the DS group from the other two groups. This can be seen graphically in Figure 4.11 below.

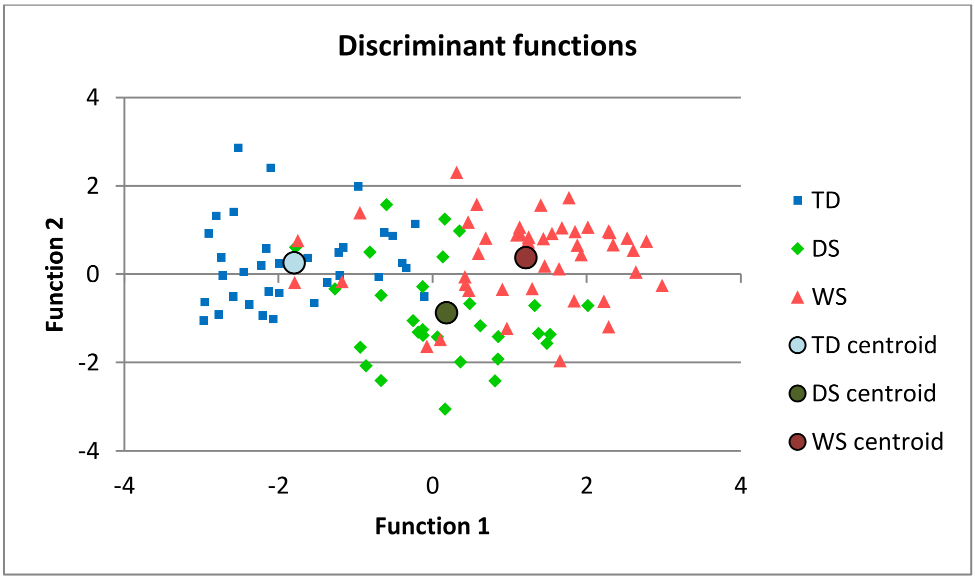

Figure 4.11: Canonical discriminant function scatterplot. Centroids show group means

Although Organization of Materials was not the only scale to account for the interaction across the three groups in the Scale by Group ANOVA, the outcome from the DFA adds weight to the pattern of results obtained from the ANOVA that saw Organization of Materials scores accounting for different patterns of performance between the atypical groups, particularly when we recall that this scale was scored relatively well within the DS group but not for the other two groups.

4.3.3.2 A three-factor approach to the BRIEF

In the BRIEF manual, the eight scales comprise two indices: the behavioural regulation index (BRI) encompassing Inhibit, Shift and Emotional Control, and the metacognition index (MI) encompassing Initiate, Working Memory, Plan/Organize, Organization of Materials and Monitor. Gioia, Isquith, Retzlaff, and Espy (2002) examined the underlying factor structure of the BRIEF in a combined clinical sample of 374 children aged between 5 and 18 years. Disorders included Attention-Deficit/Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD), Tourette's syndrome (TS), seizure disorders, affective disorders and learning disabilities. Four different models were assessed for goodness of fit using confirmatory factor analysis, and a three-factor model was chosen as the most appropriate. The factors were: Metacognition (MI), Behavioural Regulation (BR) and Emotional Regulation (ER). In this new structure, the Monitor scale was split into Task-Monitor and Self-Monitor, producing nine scales in all. The MI index includes Initiate, Working Memory, Plan/Organize, Organization of Materials and Task-Monitor; the BR index, Inhibit and Self-Monitor; and the ER index, Shift and Emotional Control. This three-factor structure was used in the current study. To compare performance across indices, mean item scores were calculated as above. Group means can be seen in Figure 4.12.

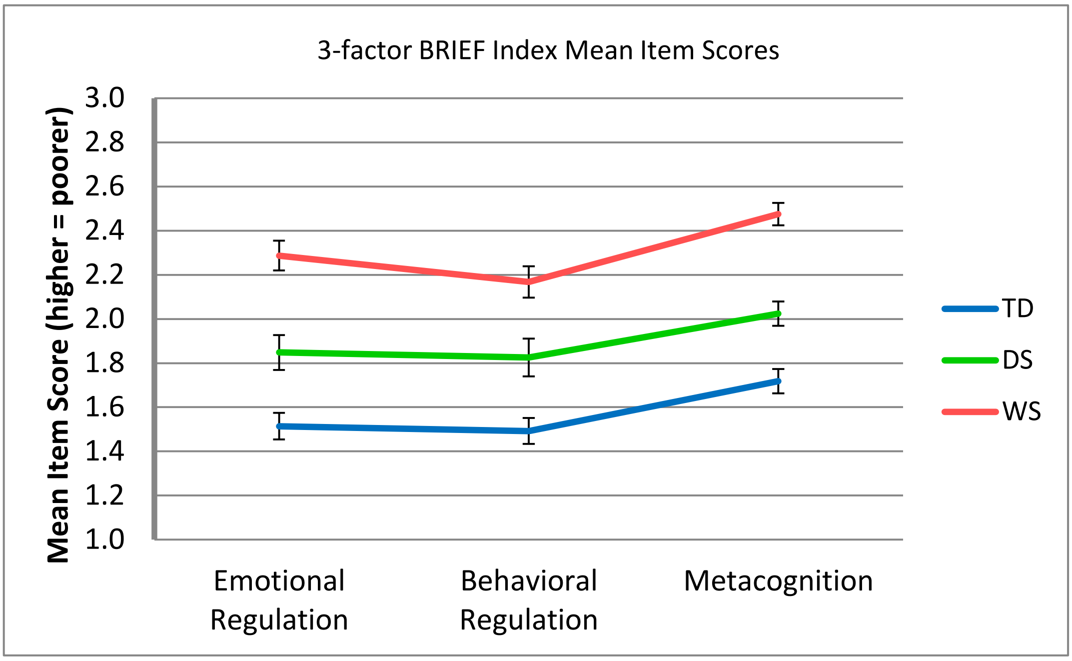

Figure 4.12: Group means *(S.E.)* of scores on three BRIEF indices by group

ANOVA revealed a significant main effect of Group ($F(2,107) = 45.028$, $p < .001$, partial $\eta^2 = .457$), with *post hoc* differences significant between all groups (TD and DS: $p = .001$; TD and WS: $p < .001$; DS and WS: $p < .001$). There was a reliable main effect of Index ($F(2,214) = 29.754$, $p < .001$, partial $\eta^2 = .218$), with statistically equivalent ER and BR scores ($p = .104$) and better scores on these scales compared to Metacognition ($p < .001$ for both). Index and Group did not interact significantly ($F<1$).

4.3.3.3 'Clinical Significance'

The BRIEF also gives scores in the form of T scores, standardised around a mean of 50 and a standard deviation of 10. The Global Executive Composite (GEC) consists of the sum of the BRI and MI indices, and gives an overall indication of executive function abilities. Gioia and colleagues (2000a, 2000b) noted that large discrepancies between the two indices (BRI and MI) that comprise the GEC would cause difficulties for its use as an overall indicator of performance: the professional manual (Gioia et al., 2000a) contains instructions not to calculate it for individuals with a BRI-MI difference of 13 T score points or more. Although it is acknowledged that any composite score of this sort could mask potentially insightful differences in more fine-grained components of a test, GEC T scores for the participants in the current study were calculated in order to gather information about the proportion of the sample that would meet the criterion of 'potential clinical significance' (a T score of 65 or above). Due to the approach taken to missing data (i.e., replacing each missing response with the mean of all given responses) the raw scores were not integers, so were rounded to the nearest whole number before reading off the T score. Of the TD group, 14.71% would be classified as having executive dysfunction of a clinically significant level; 53.33% of the DS group and 89.13% of the WS group also met this criterion. This is consistent with the group-level patterns of performance. Two participants with WS obtained a GEC raw score that was higher than the maximum provided in the table for looking up T score values, so the highest available raw score was used.

4.3.4 PSQ analysis

In the PSQ, parents chose an answer that best matched their child's behaviour, both in relation to how easy they find certain aspects of problem solving and then how likely they are to make certain kinds of responses. A copy of the questionnaire can be found in Appendix D, and a summary of questions and shared aspects, in Table 4.3 and Table 4.4.

4.3.4.1 Reaching the solution

Each participant was given a score for the 'reach the solution' part of the question. The group means of these scores are displayed in Figure 4.13.

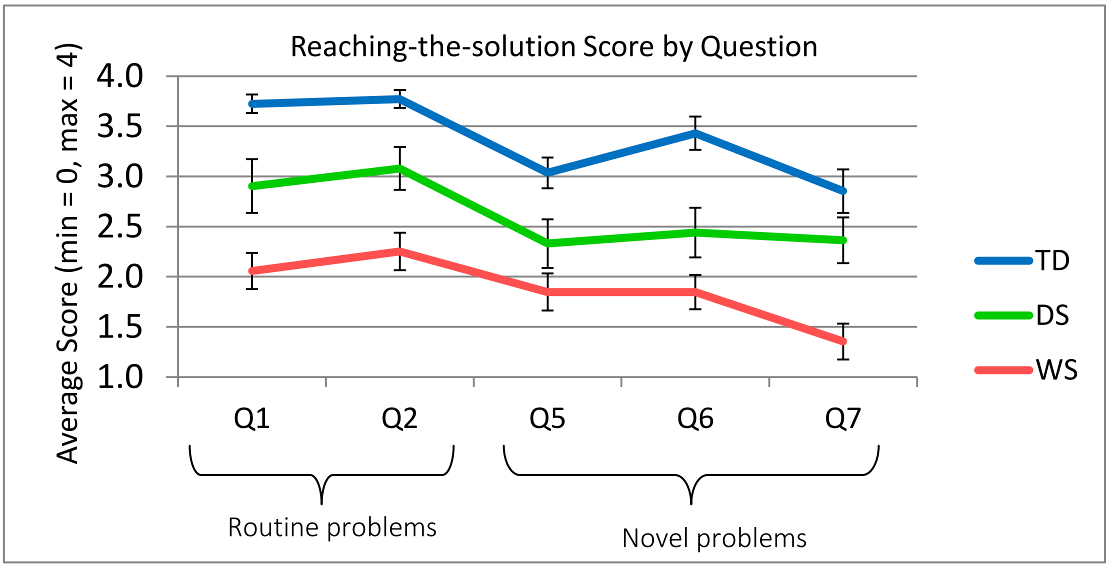

Figure 4.13: Group means *(S.E)* of reaching-the-solution score by question and group

ANOVA revealed a significant main effect of Group ($F(2,109) = 36.474$, $p < .001$, partial $\eta^2 = .401$). The pattern was similar to that for the BRIEF scores; Games-Howell *post hoc* tests indicated higher scores for TD than DS ($p = .001$) and WS groups ($p < .001$), and higher scores in DS than WS groups ($p = .001$). There was a main effect of Question ($F(4,436) = 13.504$, $p < .001$, partial $\eta^2 = .110$). Multiple comparisons did not reveal a significant difference between scores for the two routine problems, getting dressed (Q1) and brushing teeth (Q2) ($p = .257$). Each of these had significantly higher scores than each novel question ($p < .05$ for all). Within the novel questions, questions 5 and 6 did not differ from one another ($p = .155$), and question 7 was given a lower score than question 6 ($p = .003$), and marginally lower than question 5 ($p = .097$). The interaction between Question and Group was not significant ($F<1$).

The data presented for each question thus validate treating routine and novel problems differently: problems occurring within routine tasks were solved more readily than problems occurring within novel tasks, although interestingly not all novel tasks were solved with equal success.

4.3.4.2 Examining performance across the PSQ

This section explores the scores on the questionnaire. The approach taken was to treat the two halves of the questions separately because they ask about two scenarios: 'approach' is what happens when a problem exists in everyday life that needs to be addressed, and 'response' is what happens if something were to go wrong with the solving of the problem. Before going on to look at these sections separately, the overall scores are addressed briefly.

An ANOVA was conducted with Approach/response and Novelty as the within-groups variables and Group as the between-subjects variable. There was a main effect of Novelty ($F(1,109) = 94.752$, $p < .001$, partial $\eta^2 = .465$) with better scores for routine than novel problems, and a main effect of Approach/response ($F(1,109) = 99.417$, $p < .001$, partial $\eta^2 = .477$) with better scores for approach than for response. These two factors interacted significantly ($p = .047$). Given that all four differences were significant for follow-up comparisons ($p < .001$), this appears to be due to a larger drop in performance from approach to response for routine problems (t = 10.20) than for novel problems (t = 5.97). There was a significant main effect of Group ($F(2,109) = 31.725$, $p < .001$, partial $\eta^2 = .368$), which will be explored with respect to its significant interactions with other factors. There was a significant interaction between Group and Novelty ($p = .027$). Score differences between routine and novel problems (collapsed across approach/response) were significant for each group ($p < .001$ for all) but appeared to decline more for the WS group between routine and novel problems than for the other two groups. Differences between the TD and DS groups were non-significant at each level of novelty (Routine: $p = .120$; Novel: $p = .054$, although for the Novel scores the nonparametric Mann Whitney test was significant, $p = .037$). The WS group scored more poorly than both other groups at both levels of novelty ($p < .05$ for all). Approach/response also interacted significantly with Group ($p = .022$). For the approach and response scores (collapsed across novelty), the WS scores were worse than those of both other groups ($p < .05$ for all) while TD scored better than DS for approach ($p = .014$) but not for response ($p = .358$).

4.3.4.2.1 Approach

Scores for the five questions were entered into a mixed ANOVA, with Group as the between-subjects variable (TD, DS, WS) and Question (1,2,5,6,7) and Aspect as the within-subjects variables. The five Aspects used in the analysis were Know, What, Steps, Focus and Stop (full names are found in Table 4.4). These were the aspects that were common to all of the questions. Aspects that did not appear in every question are discussed later in the chapter.

There was a significant main effect of Group ($F(2,109) = 32.173$, $p < .001$, partial $\eta^2 = .371$) with the WS group scoring more poorly than the DS group ($p = .001$) and the TD group ($p < .001$), and the DS group scoring more poorly than the TD group ($p = .012$). There was a significant main effect of Question ($F(4,436) = 33.225$, $p < .001$, partial $\eta^2 = .234$) and also of Aspect ($F(2.84,309.02) = 98.337$, $p < .001$, partial $\eta^2 = .474$). Further interaction terms were significant: Question by

Group (F(8,436) = 5.021, p < .001, partial η^2 = .084), and Question by Aspect (F(9.925,.586) = 16.933, p < .001, partial η^2 = .134). All of these terms interacted with one another so that the Question by Aspect by Group interaction was significant (F(21.926,1194.967) = 2.445, p < .001, partial η^2 = .043). The Aspect by Group interaction was marginally significant (F(5.670,309.018) = 2.136, p = .053, partial η^2 = .038).

To examine the Question by Group interaction, first a Question by Group mixed ANOVA was carried out. Question approach scores were the average of each aspect score per question for each participant. Figure 4.14 displays the group means by question.

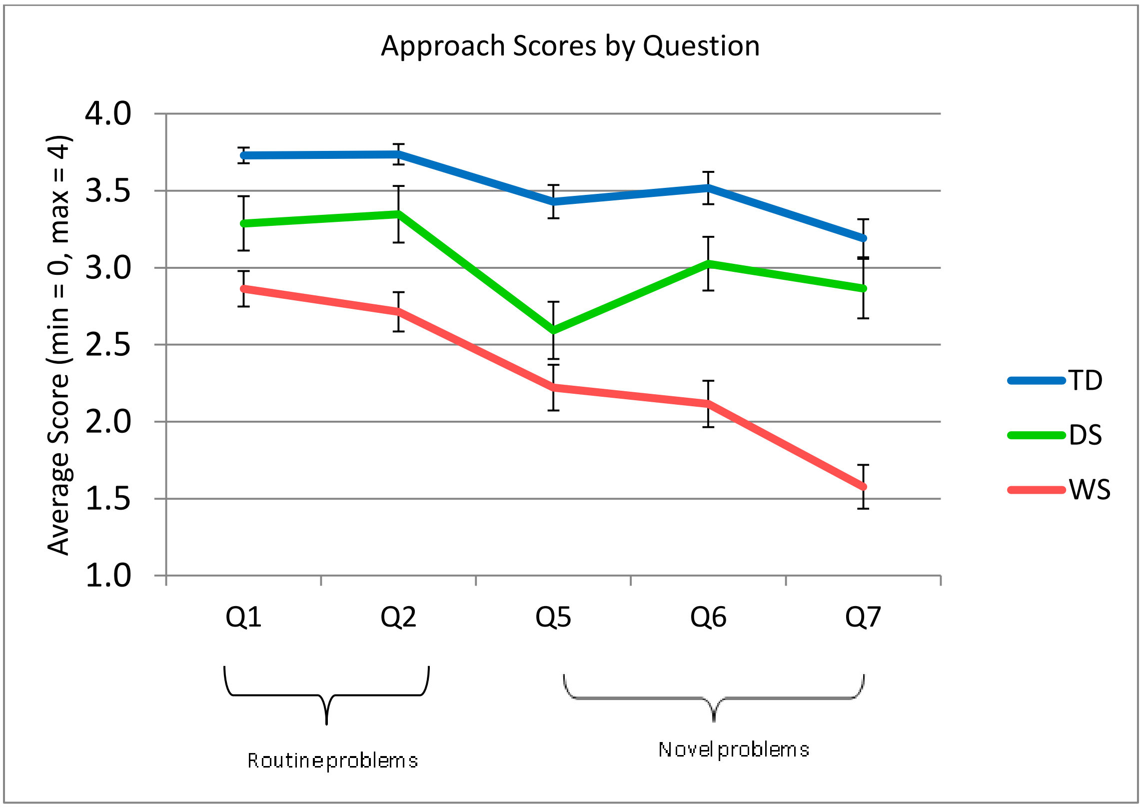

Figure 4.14: Group means *(S.E)* of shared aspect scores for approaching a problem by question

There was a significant main effect of Group (F(2,109) = 32.173, p < .001, partial η^2 = .371) with *post hoc* tests revealing poorer performance in the WS group than both of the other groups (TD: p < .001; DS: p = .001) and poorer performance in the DS group than the TD group (p = .012).

Within each group, the main effect of Question was significant (TD: ($F(3.233,106.681) = 9.567$, $p < .001$, partial $\eta^2 = .225$; DS: $F(3.200,96.002) = 6.999$, $p < .001$, partial $\eta^2 = .189$; WS: $F(3.422,157.429) = 29.670$, $p < .001$, partial $\eta^2 = .392$). *Post hoc* tests indicated somewhat similar patterns of performance by question across groups, with some notable differences. Questions 1 and 2, the routine problems, obtained the highest scores in each group. For the WS and TD groups, scores for question 7 were lower than each of the other scales ($p < .05$ for all) while for the DS group, question 5 was also scored particularly low, scoring marginally lower than question 7 ($p = .076$) and significantly more poorly than the remaining questions ($p < .05$ for all), potentially indicating particular difficulties with finding a lost possession in DS. Thus, the Question by Group interaction can be accounted for by relative difficulties with finding a lost possession within the DS group.

The 3-way interaction can be explained by the relatively poor performance on question 5 (finding a lost possession) for the DS group: *post hoc* comparisons indicated that there were no significant differences between DS and WS scores for any aspect of question 5 (although the Focus score was marginally better in the DS group, $p = .069$; using the Mann-Whitney nonparametric equivalent test, this difference was significant at $p = .021$), while the DS group outperformed the WS group on at least two aspects in all other questions (Questions 1 and 2: 2 aspects; question 6: 4 aspects; question 7: all 5 aspects). The significant interaction between Question and Aspect was also examined. Scores can be seen in Figure 4.15 below.

4.3.4.2.2 Aspect scores for each question

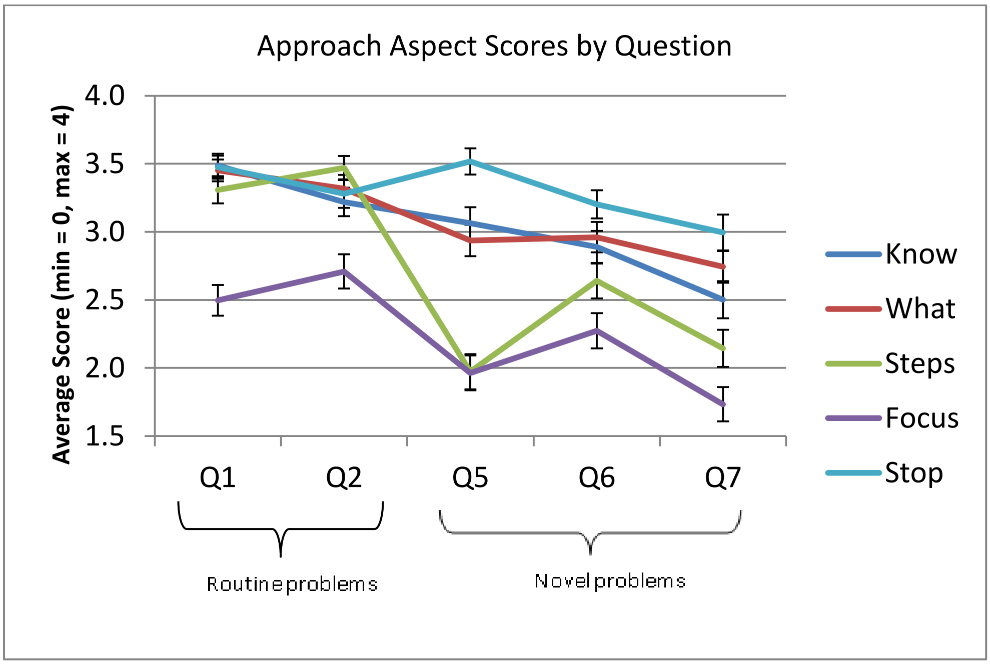

Figure 4.15: Means *(S.E)* of approach aspect scores by question, collapsed across group

As Aspect significantly interacts with Question, the scores across aspects are compared for each question separately. ANOVAs indicated a significant main effect of Aspect for each question ($p < .001$ for all). It is clear from Figure 4.15 that the Focus scores are consistently low, and this is borne out by the data, with significantly lower scores in this aspect than each other aspect in all questions ($p < .001$ for all), with the exception of question 5 in which Steps is as low as Focus ($p = .882$). Stop scores are higher than the other aspects on questions 5, 6 and 7 ($p < .05$ for all). Although it appears from examining Figure 4.15 that the Steps score is higher for routine and lower for novel questions, the only question in which Steps is higher than other aspects is question 2 (brushing teeth), in which it is significantly higher than Know ($p = .009$) and marginally higher than What ($p = .071$) and Stop ($p = .061$). In contrast, in question 1 Steps is significantly poorer than Know ($p = .021$), What ($p = .035$) and Stop ($p = .037$), while for the novel questions Steps is poorer still than Know, What and Stop ($p < .001$ for all, except Know, Q6: $p = .009$). Therefore, 'doing steps in the right order' was particularly successful when brushing teeth, and showed a larger drop in performance, relative to other aspects, for the novel problems than for the other routine problem.

Finally, there are some aspects which did not appear in every question. Recognising the problem was only part of the Approach part of the question for novel problems, and questions about using existing knowledge and applying a previously learnt strategy were unique to novel problems. These scores were averaged across question for each participant and compared across group. For all three of the aspects the effect of Group was significant at $p < .001$, and the same *post hoc* pattern of performance held, mirroring overall group performance: the TD group outperformed the DS group (Recognise $p = .004$; Existing Knowledge $p = .001$; Applying strategy $p < .001$) and the DS group outperformed the WS group, although marginally so for the Recognise aspect, although this is significant for the nonparametric analysis (Recognise $p = .076$, Mann-Whitney $p = .023$; Existing Knowledge $p = .016$; Applying strategy $p = .004$). In addition, two questions included a unique item, pertinent to that scenario. Question 5 asks about ending the search after a sensible amount of time. ANOVA indicated a main effect of Group ($F(2,109) = 4.159$, $p = .018$, partial $\eta^2 = .071$). *Post hoc* tests revealed equivalent scores between DS and TD groups ($p = .131$); equivalent scores between DS and WS groups ($p = .811$); and significantly better performance in the TD group compared to the WS group ($p = .009$). Question 6 asks about children's success with being flexible when packing a bag for the day. ANOVA indicated a main effect of Group ($F(2,109) = 13.707$, $p < .001$, partial $\eta^2 = .201$). The TD group scored better than the DS group ($p = .022$) and WS group ($p < .001$), while there was no significant difference between the atypical groups ($p = .109$), although in a Mann Whitney test the DS score was significantly higher than the WS score ($p = .048$).

4.3.4.2.3 Response

Response scores for each shared aspect were entered into an Aspect (Change, Ask, Become Emotional, Lose Focus and Lack Perseverance) by Question (1, 2, 5, 6, 7) by Group (TD, DS, WS) ANOVA. There was a significant main effect of Group ($F(2,109) = 24.334$, $p < .001$, partial $\eta^2 = .309$), with *post hoc* tests revealing better performance in the TD than the WS group ($p < .001$) and better performance in the DS than the WS group ($p < .001$) but no significant difference between TD and DS groups ($p = .369$).

The main effect of Question was significant ($F(3.260,355.316) = 21.751$, $p < .001$, partial $\eta^2 = .166$) and did not interact with Group ($F(6.520,355.316) = 1.393$, $p = .211$, partial $\eta^2 = .025$). The effect of Aspect was also significant ($F(2.717,296.168) = 4.727$, $p = .004$, partial $\eta^2 = .042$), as was the Aspect by Group interaction ($F(5.434,296.168) = 7.864$, $p < .001$, partial $\eta^2 = .126$): thus, the main effects of Aspect will be explored in relation to Group. Question and Aspect interacted significantly ($F(10.148,1106.092) = 25.125$, $p < .001$, partial $\eta^2 = .187$), so the main effect of Question will be explored in relation to Aspect. Finally, the interaction between Question, Aspect and Group was also significant ($F(20.295,1106.092) = 2.168$, $p = .002$, partial $\eta^2 = .038$).

Across questions, the shared aspects were compared across groups using repeated-measures ANOVA. Figure 4.16 displays the mean scores by aspect for each group. The reach-the-solution score is not part of the current analysis but a data point has been added to illustrate the mean score on this measure for each group.

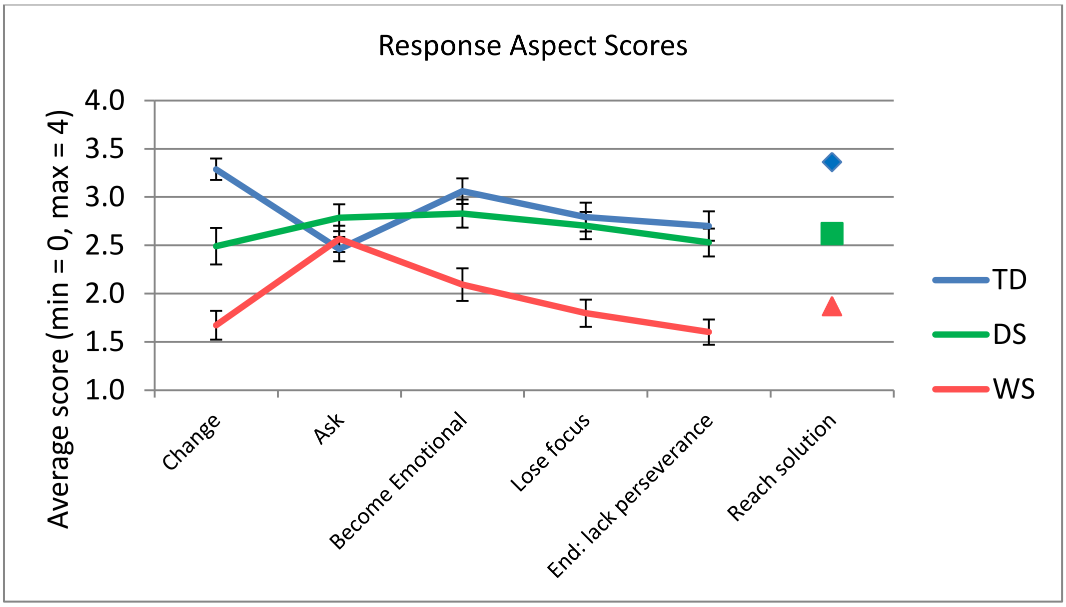

Figure 4.16: Group means *(S.E)* of response aspect scores

The Aspect by Group interaction was driven by a lack of Group effect for Asking for help ($F(2,109) = 1.303$, $p = .276$, partial $\eta^2 = .023$) while all other aspects showed a main effect of Group (Change: $F(2,109) = 29.360$, $p < .001$, partial $\eta^2 = .350$; Become Emotional: $F(2,109) = 11.223$, $p < .001$, partial $\eta^2 = .171$; Lose focus: $F(2,109) = 15.669$, $p < .001$, partial $\eta^2 = .223$; Lack perseverance: $F(2,109) = 18.959$, $p < .001$, partial $\eta^2 = .258$). *Post hoc* tests revealed better scores in TD than WS for each of the remaining four aspects ($p < .001$ for all). For the Change score, the TD group scored higher than the DS group ($p = .002$) and the DS group scored higher than the WS group ($p = .004$). For the remaining three variables (Become emotional, Lose focus, Lack perseverance) performance was equivalent in the TD and DS groups ($p > .05$ for all) and better in the DS than the WS group (Become emotional: $p = .004$; Lose focus: $p < .001$; Lack perseverance: $p < .001$).

There were also some patterns of note in the within-groups analysis. Repeated-measures ANOVAs indicated main effects of Aspect for TD ($F(2.072,68.373) = 7.737$, $p = .001$, partial $\eta^2 = .190$) and WS groups ($F(2.396,110.196) = 11.574$, $p < .001$, partial $\eta^2 = .201$) but not for the DS group ($F(3.242,97.263) = 1.633$, $p = .183$, partial $\eta^2 = .052$) although Friedman's (nonparametric) ANOVA for the DS group was significant ($\chi^2(4) = 11.719$, $p = .020$). While the TD group were more likely to change their response than ask for help ($p < .001$), the atypical groups were both more likely to ask for help than to change their response to a problem; this was a stronger effect in WS than in DS (DS: $p = .070$; WS: $p < .001$). Aspect scores displayed by question (collapsed across Group) are displayed in Figure 4.17.

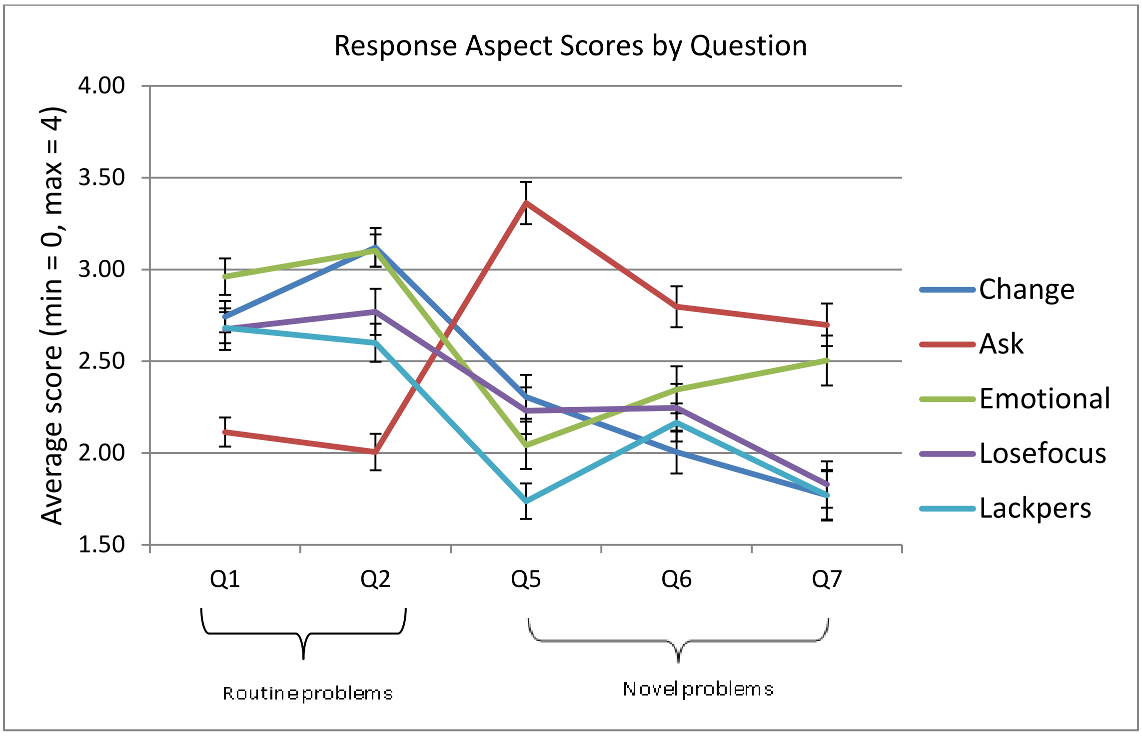

Figure 4.17: Means *(S.E)* of response aspect scores by question, collapsed across group. Lackpers = lack perseverance.

It is clear from examining Figure 4.17 that the pattern of scores for the Ask for Help variable is different across questions compared to the other aspects. ANOVAs reveal that, indeed, Ask for Help scores are lower than scores for all other aspects in questions 1 and 2, and higher than scores for all other aspects in questions 5, 6 and 7 (p < .05 for all) with the exception of the non-significant difference between Ask for help and Become Emotional in question 7 (p = .366). Asking for help occurred most often in question 5, then in questions 6 and 7, and least often in questions 1 and 2, with equivalent scores in questions 6 and 7 (p = .409), and 1 and 2 (p = .463), and all other question differences significant for this aspect (p < .001). Thus, broadly speaking, asking for help occurred more for novel than routine problems.

The Aspect by Group interaction was significant for questions 1 and 2 (p < .001 for both), 6 (p = .001) and 7 (p = .020), but non-significant for question 5 (p = .106). Although for the Approach analysis, finding a lost possession (question 5) showed a different pattern of performance to the other questions with regard to aspect (i.e., with DS aspect scores being as low as the WS scores for each aspect on this question but not the other questions), for the Response part of question 5 this was not the case.

Based on observation of the data, the 3-way interaction was explored by determining the effect of Group for each aspect for each question. One-way ANOVAs point to significant main effects of Group in terms of Change, Emotion, Lose focus and Lack perseverance for each question ($p < .05$ for all), as well as a lack of reliable main effects of Group for Asking for Help for questions 1, 5, 6 and 7 ($p > .05$ for all). However, there was a significant main effect of group for asking for help for question 2, ($F(2,109) = 8.628$, $p < .001$, partial $\eta^2 = .137$) with *post hoc* tests revealing that the TD group were less likely to ask for help when brushing teeth than the WS ($p = .007$) or DS group ($p = .001$) with no significant differences between the atypical groups ($p = .357$).

4.3.4.3 Which variables are related to reaching the solution?

4.3.4.3.1 Approach variables

In order to assess the relationship between success with different aspects of approaching a problem and being able to solve it, correlations were run between each of the shared approach aspects and the reaching-the-solution score, for each question individually. The pattern of association between success and reaching the solution applied in a broadly similar way across the usable questions (1, 2, 5, 6 and 7) for four of the aspects (Know, What/Ideas, Focus, Stop). Thus, scores were collapsed across question for these aspects. The correlations with reaching the solution are displayed in Table 4.6.

R (*p*)	TD	DS	WS
Know	.623 (<.001)	.738 (<.001)	.642 (<.001)
What	.640 (<.001)	.714 (<.001)	.703 (<.001)
Steps	.749 (<.001)	.833 (<.001)	.730 (<.001)
Focus	.577 (<.001)	.817 (<.001)	.778 (<.001)
Stop	.537 (.001)	.757 (<.001)	.447 (.002)

Table 4.6: Correlations (r, with *p* value in parentheses) between PSQ approach aspects and reaching the solution by group. Significant correlations are displayed against a green background.

For the Steps variable, the significant relationship with reaching the solution only held in the WS group for question 2 (r = .388, $p = .007$) and for the DS group, for questions 1 and 2 (Q1: r = .645, $p < .001$; Q2: r = .683, $p < .001$), with none of the questions giving this relationship in the TD group ($p > .05$ for all).

4.3.4.3.2 Response variables

Correlations were also conducted between reaching the solution and each of the other behaviours in the response part of the question, collapsed across questions in the same way as for the approach part of the question, because the patterns of relationships were again broadly similar for each question. It should be noted that the correlations are positive because the scores were reversed for the Become Emotional, Lose focus and Lack perseverance scales.

R (p)	TD	DS	WS
Change	.575 (<.001)	.897 (<.001)	.746 (<.001)
Ask	-.107 (.549)	.605 (<.001)	.133 (.374)
Become Emotional	.259 (.139)	.277 (.131)	.449 (.002)
Lose focus	.521 (.002)	.436 (.014)	.609 (<.001)
Lack perseverance	.538 (.001)	.556 (.001)	.613 (<.001)

Table 4.7: Correlations (r, with p value in parentheses) between PSQ response aspects and reaching the solution by group. Significant correlations are displayed against a green background.

4.3.4.3.3 Additional variables

Some aspects do not appear in every question and so the correlations were conducted separately. Separate correlations were conducted for the 'recognise that there is a problem' aspect because it appeared in the approach part of the question for novel problems, but the response part for routine problems. Across novel problems, recognising the problem was significantly correlated with reaching the solution for all groups (TD: r = .541, p = .001; DS: r = .659, p < .001; WS: r = .635, p < .001) and the same was true for routine problems (TD: r = .503, p = .002; DS: r = .757, p < .001; WS: r = .338, p = .020).

In the approach part of each novel problem, parents are asked about their child's ability to use existing knowledge to help them, and to apply a previously used strategy. The mean aspect score across novel questions was correlated with the Reach the solution score across novel questions: each correlation was highly significant. For using existing knowledge, TD: r = .741, p < .001; DS: .666, p < .001; WS: r = .724, p < .001; for applying a previously learnt strategy, TD: r = .573, p < .001; DS: r = .725, p < .001; WS r = .679, p < .001.

Finally, some aspects only appear in one question. In question 6 (packing a bag for the day) parents are asked how easy their child finds it to be flexible when packing a bag (approach part of the question). Both TD and WS correlations are significant (TD: r = .492, p = .003; WS: r = .555, p < .001) and the DS correlation is marginally so (r = .353, p = .051). In question 5 (finding a lost possession) parents are asked how likely their child is to end their search after a sensible amount of time (response part of the question). In TD and WS, again, this is reliably associated with reaching the solution (TD: r = .557, p = .001; WS: r = .451, p = .001) but this is non-significant in DS (r = -.009, p = .960).

4.3.4.3.4 BRIEF scale scores

Given that reaching the solution has been treated as the main indicator of problem-solving success, correlations were run between each BRIEF scale score and reaching the solution, collapsed across question for routine and novel problems separately, in order to identify any BRIEF scales that were associated with success on different types of everyday problems. Pearson's correlation coefficients, r, and p values can be seen in Table 4.8. Significant correlations ($p \leq .05$) are against a green background; the nonparametric Spearman's correlation outcome is also reported where the significance of the result differs from that produced by the parametric correlation. Marginally significant correlations ($.05 < p < .10$) have a yellow background. Table 4.8 displays correlations for the eight BRIEF scales from the manual. The factor structure used in Section 4.3.3.2 was based on nine scales, due to the splitting of the Monitor scale into Task-Monitor and Self-Monitor. Thus, correlations for these two new scales are also presented in Table 4.9.

	TD		DS		WS	
	Routine	Novel	Routine	Novel	Routine	Novel
	R (*p*)					
Inhibit	-.092 (.605)	-.439 (.009)	-.414 (.023) -.316 (.089)	-.585 (.001)	-.219 (.145)	-.431 (.003)
Shift	.037 (.835)	.035 (.845)	-.494 (.006)	-.318 (.087)	-.297 (.045) -.268 (.072)	-.492 (.001)
Emotional Control	.082 (.645)	-.070 (.692)	-.460 (.010)	-.577 (.001)	-.256 (.085)	-.474 (.001)
Working Memory	-.299 (.086)	-.303 (.082)	-.549 (.002)	-.700 (<.001)	-.153 (.312)	-.581 (<.001)
Plan/Organize	.031 (.862)	-.233 (.185)	.167 (.377)	-.590 (.001)	-.232 (.120)	-.330 (.025) -.224 (.135)
Organization of Materials	-.102 (.565)	-.226 (.199)	-.292 (.117)	-.714 (<.001)	-.155 (.305)	-.562 (<.001)
Initiate	.003 (.987)	-.039 (.826)	-.404 (.027)	-.428 (.018)	-.356 (.015)	-.504 (<.001)
Monitor	-.171 (.334)	-.319 (.066)	-.352 (.056)	-.458 (.011)	-.142 (.346)	-.425 (.003)

Table 4.8: Correlations (r, with p value in parentheses) between BRIEF scale scores and reaching-the-solution score on the PSQ for routine and novel problems. Significant correlations are displayed against a green background, while marginally significant correlations are against a yellow background.

The 48 correlations in Table 4.8 would require a *p* value cut-off of 0.001042 in order to survive a Bonferroni correction. For routine problems, none of the correlations would survive this correction. For novel problems, the surviving correlations between reaching the solution on the PSQ and the BRIEF scales for the DS group would be for the Inhibit, Emotional Control, Working Memory, Plan/Organize and Organization of Materials scales. For the WS group, surviving correlations between would be for the Shift, Emotional Control, Working Memory, Organization of Materials and Initiate scales. None of the TD group's correlations would survive a Bonferroni correction.

	TD		DS		WS	
	Routine	Novel	Routine	Novel	Routine	Novel
	R (*p*)					
Task-Monitor	-.131 (.461)	-.381 (.026)	-.208 (.271)	-.345 (.062)	-.199 (.185)	-.339 (.021)
Self-Monitor	-.135 (.447)	-.115 (.516)	-.390 (.033)	-.442 (.014)	-.063 (.676)	-.392 (.007)

Table 4.9: Correlations (r, with p value in parentheses) between Task-Monitor and Self-Monitor scales and reaching-the-solution score on the PSQ for routine and novel problems. These two scales combined comprise the Monitor scale in Table 4.8. Significant correlations are displayed against a green background, while marginally significant correlations are against a yellow background.

In Table 4.9, 12 correlations were conducted, requiring a *p* value cut-off of 0.004167 to survive a Bonferroni correction; none of the correlations would survive this correction.

4.3.4.3.5 Chronological age

Correlations between CA in each group and reaching the solution for routine and for novel problems were also conducted. For the TD group, neither correlation was significant (Routine: r = .273, p = .118; Novel: r = .261, p = .135) and this lack of association was also seen for the WS group (Routine: r = .193, p = .193; Novel: r = .101, p = .497). However there was a significant association between CA and reaching the solution in novel problems for the DS group (r = .384, p = .033) while for routine problems this was not the case (r = .134, p = .472).

4.4 Discussion

Parents completed two questionnaires about their children's everyday EF and problem solving: the BRIEF and the PSQ.

4.4.1 Group performance across questionnaires

The same performance pattern across the groups was obtained on both questionnaires; that is, better performance in the typical than the atypical groups, and better performance in the DS group than the WS group (for further discussion see Section 4.4.5). This implies that the PSQ, designed for this study, taps into everyday functioning in a similar way to the BRIEF, which is a standardised and validated measure, thus adding credibility to the PSQ overall score as an indication of everyday functioning, even though the questions are posed from a different angle to those in the BRIEF.

The overall scores for each questionnaire indicated that on the BRIEF, the vast majority of the WS group presented with performance that would indicate potential clinical significance in everyday EF, and approximately half the DS group fell into this category, with a much lower percentage in the TD group. These outcomes are notable in two senses, the first being that both atypical groups demonstrated deficits in everyday functioning, relative not only to their CA but to a group of younger children. Given that the mean CA of each atypical group was above 17 years, and the mean CA of the TD group was just over 8 years, this indicates a substantial impairment in the WS and DS groups' performance as both groups, on average, performed even more poorly than children around nine years younger. The second sense in which the group patterns in results are of note is that one disorder (WS) demonstrated reliably poorer performance than the other disorder (DS) though the groups' CA was not significantly different. This illustrates, in a simple way, the importance of engaging in research across syndromes: rather than simply finding poorer performance than TD children in a neurodevelopmental disorder, we can observe levels of ability that are syndrome-specific.

Significant associations between CA and percentage PSQ (and BRIEF) scores on both questionnaires were seen for WS and DS groups in Section 4.3.2.4. This is echoed for novel (but not for routine) problems for the DS group regarding reaching the solution on the PSQ, being only significantly associated with increasing CA for novel problems for the DS group, with TD and WS groups showing no association between CA and reaching the solution on either routine or novel problems. Thus while it seems that some aspects of everyday problem solving were developing with CA for the WS group, these were not impacting their ability to actually solve the problem. Furthermore, some discrepancies were seen between the presence of correlations between TOL and CA (for the TD group only) and the questionnaire percentage scores and CA (for the atypical groups only). Without knowing what the relationships would be between these measures for the full range of CAs (i.e., from the lowest TD CA to the highest WS/DS CA in all the populations) it is difficult to speculate on the reasons for this. However, for the WS group only BPVS score was related to CA (Chapter 2), while for the DS group there were age-related improvements on a small number of EF/MA measures (planning, inhibition accuracy, BPVS and RCPM). Speculatively, if the individuals with WS were acquiring general problem-solving skills in everyday life which nevertheless did not enable them to solve the real-life problem or improve on EF tasks in controlled settings; but the DS group were improving on skills more generally, which *were* impacting their everyday problem-solving success and their abilities on some EF tasks, then the current pattern of results would be produced. Further research should address these tentative hypotheses.

4.4.1.1 Group comparisons

The WS group had a verbal ability (BPVS score) that was not reliably different from that of the TD group, and was better than that of the DS group. Nonetheless, WS questionnaire outcomes were significantly worse than those of each of the other groups, emphasising the marked uneven nature of the WS cognitive profile, as described by, for example Mervis et al. (2000), and the discrepancy between verbal scores and problem solving in the WS group. As the WS group did not score more poorly than the DS group on either mental age measure and yet still performed more poorly on the questionnaire measures, mental age differences between the atypical groups did not account for differences in performance. This type of conclusion can of course only be made in relation to the particular matching measure that is chosen.

4.4.1.2 BPVS

BPVS scores were related to both questionnaire scores in the WS group, and only marginally to the PSQ in the TD group. The significant relationship between BPVS and questionnaire scores in the WS group suggests that increasing verbal ability facilitated better problem-solving skills for this group. Trajectory analyses of the questionnaire scores in relation to BPVS score revealed that the TD and DS groups' scores did not differ either at the lowest levels of verbal ability, or throughout increasing levels of verbal ability (i.e., there was no reliable difference in slope between the groups), indicating that the difference in TD and DS group means on the PSQ was not present when verbal ability was taken into account. The WS group performed more poorly on the PSQ at the lowest levels of BPVS score than both of the other groups (driving the overall group difference at the intercept) and as the slopes did not differ between any of the pairs of groups, indicating that BPVS and PSQ were related in a uniform way across all three groups, this points to poorer PSQ performance in the WS group regardless of verbal ability.

4.4.1.3 RCPM

The trajectory analyses for the TD and DS groups indicated better TD than DS questionnaire scores at the lowest RCPM scores. The slopes did not differ, indicating a constant relationship between questionnaire outcomes and nonverbal mental age (although for the BRIEF, the interaction was marginal, tentatively indicating a dynamic relationship: as nonverbal ability increased, the DS scores converged onto the TD trajectory).

The WS group's slope differed significantly from that of both of the other groups, indicating atypical differential relationships between RCPM score and outcomes on both questionnaires. At the lowest level of RCPM ability, the WS scores were lower than that of TD individuals but not lower than that of DS individuals. However, the WS trajectory diverged from both other groups', with decreasing questionnaire outcomes associated with increased RCPM scores. The pattern with respect to the WS group was unexpected and seems to suggest that those individuals with stronger nonverbal ability are less able to cope with problem solving in the real world than those who do worse on the RCPM. It is likely that a mediating factor played a role in producing this result (e.g., parent expectations or anxiety in WS), but as this pattern was not also observed in relation to BPVS scores, it is difficult to identify.

4.4.1.4 Relationships between questionnaires

In all of the groups, a greater number of significant correlations between BRIEF scale scores and reaching-the-solution scores on the PSQ were seen for novel problems than for routine problems, supporting the general consensus that EFs are more drawn upon when problems are more novel. More BRIEF scales were related to reaching the solution in the atypical groups than in the TD group. This might reflect the better scores in the typical group, and therefore the relative ease that TD children had in reaching the solution to these problems. The only scales related to reaching the solution in this group were Inhibit and Task-Monitor, suggesting that success was only dependent on their avoiding distractions and keeping focus on the task. In contrast, more EFs may be related to reaching the solution in the WS and DS groups if they need to use more effortful resources to be able to reach the solution.

4.4.2 Performance on the BRIEF: indices and scales

In the BRIEF manual, Gioia et al. (2000a) note (citing Barkley, 1996, 1997) that the type of problem solving delineated on the MI is likely to build on the characteristics seen on the BRI: that is, that it is probable that behavioural regulation underlies metacognition. Woodruff-Borden et al. (2010) also found higher (i.e., poorer) MI than BRI scores in a sample of children with WS, as was found here for all of the groups.

While the current approach was to analyse scales in terms of three and not two indices, for each group, the behavioural regulation and emotional regulation indices were scored better than the metacognition index. Thus, it is possible that small difficulties with self-regulation could manifest more strongly as difficulties with metacognitive EFs (e.g., planning), in line with Gioia and colleagues' suggestion. However, Lee et al. (2011) suggested the opposite pattern for DS. They found worse scores compared to MA for the Emergent Metacognition index, but not for the Inhibitory Self-Control or Flexibility indices (the preschool version of the BRI), in a group of children with DS. Citing Capone et al. (2006)'s evidence for difficulties with rigid behaviours and emotional control later in life, they suggest that these behaviour patterns could develop as a downstream effect of the early working memory and planning/organisation deficits that they found. Lee et al. (2011) added a layer of interpretation to the BRIEF indices of 'hot' EFs (i.e., more emotion-related) being represented by the Inhibitory Self-Control and Flexibility indices (later, BRI) and the 'cool' EFs (i.e., more based on cognitive processes) being represented by the metacognition index.

It is not possible to determine a direction of causality between hot and cool EF development from the current pattern of data. In addition, the finding of better self-regulation (ER; BR) than metacognition (MI) scores applied to all groups in the current study. However, rigid behaviours in DS (Capone, Goyal, Ares, & Lannigan, 2006) could be reflected in the poorer Shift scores in the DS than TD group, but equivalent Organization of Materials scores in the DS and TD groups.

Interestingly, the discriminant functions analysis (DFA) indicated that it is the function on which only the Organization of Materials scale loaded onto that differentiated the DS group from the other two groups. Skill at organising one's belongings and not leaving disorganised collections of possessions lying around is seen as positive in general everyday life as well as within the paradigm of the BRIEF. This is wholly understandable, given that, anecdotally, children are not known for readily being tidy, but that this is something that they need to be taught (or may choose not to do in order to assert their independence; Karmiloff-Smith, personal communication, 2013). This could be what is reflected by the relatively poor scores in the TD group, compared to all of the other scales. In DS, however, it could be that difficulties with shifting (Lanfranchi et al., 2010) and flexible behaviour (Capone et al., 2006) lead to a compensatory strategy of becoming very rigid with one's possessions, and liking things to be kept in order. This would then produce a score on a measure of everyday life like the BRIEF that looks like a strength, and indeed may be experienced as such in everyday life, while, in an experimental setting, the underlying difficulties with shifting may be tapped into.

If the Organization of Materials scale is something that improves in TD and emerges gradually from difficulties with shifting in DS, we might expect to see these abilities changing with CA. In fact, CA was not correlated with the Organization of Materials scale in any of the groups. One explanation for this could be that parent expectations might modulate this: absolute levels of performance on the scale could be increasing in line with parent expectations of the child, meaning that the rating of the child's ability would remain constant with age. Alternatively if these abilities do develop with age in either of the groups, it may be that this had occurred at younger ages than those of the sample, which is most likely for the DS group, being older than the TD group. RCPM scores were weakly related to this scale in the WS and TD groups only. It is conceivable that nonverbal ability might be related to being able to organise objects ordinarily, and given that the Organization of Materials scale differentiates the DS group from the other two, that something different is happening in DS: perhaps, that rather than being driven by nonverbal ability, something else is pushing the development of this skill in DS. However this must remain a tentative suggestion, due to the marginal correlations between RCPM and Organization of Materials in the WS and TD groups.

4.4.3 Performance on the PSQ

4.4.3.1 Reaching the solution

On the PSQ, the reaching-the-solution scores produced the same group pattern of results as did the overall percentage score: the TD group were more likely to reach the solution overall than the DS group, and the DS group were more likely to reach the solution than the WS group. Reaching the solution was also more likely to occur for routine problems than for novel problems, and similar patterns of performance across question were found for each group.

4.4.3.2 Overall PSQ analysis

Regarding the overall PSQ scores, routine problems were again solved more successfully than novel problems, and this was the case for each group. This is in line with expectations: routine, overlearned tasks are easier than tasks that are unusual and unexpected. It is also in line with previous research and models suggesting that greater EF resources will be drawn on when tasks are novel (e.g., Shallice & Burgess, 1991b; Stuss, 1992).

There was also an interaction between novelty and group. While the WS group scored more poorly than both other groups for routine and for novel problems, the DS group only scored more poorly than the TD group for novel problems (using the nonparametric test, due to the Likert-type questionnaire data). Therefore, the novel problems may have been disproportionately harder than the routine problems for the DS group than the TD group.

One phenomenon that might account for the better scores in the DS group than the WS group on both questionnaires is described by Vicari et al. (2000, 2001).

Compared to MA-matched controls, children with DS did not show deficits in procedural learning (implicit memory) tasks (in fact, on the TOL task), while children with WS did. Implicit memory is reflected in the ability to learn from previous experience with a task, when one is not aware of doing so. If individuals with DS are more able to learn from previous task experience than individuals with WS, it would follow that they would be in a position to gain an advantage over the WS group in tasks that are repeated very often which is available to a lesser extent in the WS group. Better scores on everyday skills in general could therefore be expected in the DS than WS group, which is reflected in the results of the current study. However, more specifically, this might predict more of a DS advantage over the WS group for the routine problems on the questionnaire than for the novel problems, because routine problems are by definition encountered more frequently, and this was not borne out by the data; therefore, there are probably additional factors at play.

Approach scores were higher than response scores, indicating that carrying out a task is easier before something goes wrong than afterwards. This is intuitive, as even the novel problems on the questionnaire were necessarily situations which had been previously encountered in order to be able to answer the question. Arguably the true problem-solving element of the situation would be called into play in the response part of the question, where perhaps more opportunities for failure arise. This effect was exaggerated in the TD group, bringing TD scores in line with DS scores overall at the response stage, and accounting for the interaction between approach/response and group. This could suggest a disproportionate drop in competence for TD children when faced with something going wrong with a task which is not seen in the other groups, potentially reflecting the TD group's younger age.

The interaction between approach/response and novelty demonstrates a decline in performance from approach to response that (while all differences were significant) looked larger for routine problems than for novel problems (indeed, the routine problems t-test statistic was larger: see Section 4.3.4.2). Responding to something going wrong seems to introduce more difficulty for routine than for novel problems, potentially due to the initial good level of performance for the relatively easy task of approaching a routine problem.

4.4.3.3 PSQ: Approach section

For the approach part of the PSQ, the same group pattern of performance was reflected as for reaching the solution: better scores in TD than DS groups, and better scores in DS than WS groups. This was broadly mirrored for the aspects that did not appear in every question, with the exception of ending one's search after a sensible amount of time (question 5) for which the group difference did not reach significance between the DS group and either of the other groups.

Across questions, the Focus aspect was scored consistently poorly, indicating that overcoming distractions was the most difficult part of dealing with the tasks in everyday life. This is seemingly at odds with the results of the BRIEF, which indicate that the Inhibit scale attracted the best score for the WS and DS groups. This can be reconciled when the items that comprise the Inhibit score on the BRIEF are examined. They appear to be concerned with social aspects of inhibition (see Appendix C), while the Focus aspect on the PSQ is concerned with successfully ignoring distractions.

Question 5 (finding a lost possession) stood out from the other questions for two reasons. First, Steps/Keeptrack was equally as poor as the Focus aspect for this question. This is likely to be due to the nature of the finding a lost possession task: unlike the other four questions (getting dressed, brushing teeth, packing a bag for the day, putting things away in a chest of drawers/wardrobe) this task perhaps is the least predictable and has the least amount of structure. Keeping track of the task involves remembering where you have already looked and is probably more difficult than keeping track when, say, packing a bag, when it is easy to look inside the bag to monitor progress, which might serve to support motivation. Encouragement from others could also be more easily delivered when it is clear that a step has been taken (e.g., an item packed) and the individual can be directed to the next small step. Second, the DS group scored poorly on question 5, both scoring similarly to the WS group for each aspect and scoring relatively poorly on this task in relation to other tasks. The reasons above might account for why this task might be difficult, but do not readily suggest why the DS group found this task particularly difficult compared to other tasks whereas the WS group did not. One explanation for this might be that procedural, or implicit, learning (discussed above) may have facilitated performance in the DS group on tasks that have certain structures and cues to their progression, such as packing a bag for the day or putting items away in a wardrobe, while implicit learning may be less helpful in a situation where, because the item is consistently not found until it is found, the steps along the way to solving the problem are less salient. An alternative explanation might be rooted in personality profiles: it may be that the lack of salient structure of the finding a lost possession task makes it intrinsically less motivating during task solving than something on which the steps of progress can be monitored. Individuals with WS are known to be strongly motivated by social factors, even being labelled as 'hypersocial' (Jones et al., 2000). It is therefore possible that motivation in this group could be maintained by general encouragement from (for example) a parent, while individuals with DS may require more tangible motivation such as that which is available from measurable progress, meaning that when this is not available, performance would decline. In addition, the poor performance when an item is lost might be linked to a liking for keeping things tidy, implied by the relatively good Organization of Materials BRIEF scale score for the DS group. Perhaps the individuals with DS liked to keep things tidy (scoring highly on Organization of Materials) because they found losing an item particularly aversive.

For question 7 (putting items away in a chest of drawers/wardrobe), the WS and TD groups scored more poorly than for each other question. Poor performance on this task is in line with expectations that were set when speaking to parents about everyday problem solving in WS. Anecdotally, keeping rooms tidy and in order is difficult for individuals with WS. The Organization of Materials score would also speak to this type of ability, and this scale was among the poorer scales for the WS group. Interestingly, (as referred to in the Introduction) it was suggested to the author by a practitioner that brushing teeth (question 2) causes difficulties for individuals with DS. This was not borne out by the data, but this could be because it was compared to scenarios which are likely to be more difficult (i.e., the novel problems). However, unusual response patterns were observed for question 2 overall, in which the Steps aspect was stronger than most other aspects, whereas for the other questions it was relatively poor. This might be an artefact of the design of the questionnaire: brushing one's teeth is perhaps the most narrow and well-practised task of them all, with the least amount of room for deviation from a well-practised path, leading to good Steps scores.

4.4.3.4 PSQ: Response section

For response scores, the WS performance was poorer than that of the other groups but the TD and DS performance did not significantly differ, because of the more dramatic drop in scores between approach and response for the TD group (discussed above). Asking for help was more likely to happen in novel problems than in routine problems. This is intuitively appealing, given that novel problems are more difficult than routine problems.

When responding to a problem, the TD group was more likely to change their response than the DS group, who were more likely to do this than the WS group. In contrast, asking for help was equally likely for all three groups, while the remaining aspects were scored more poorly in the WS than in the DS or TD groups, who did not differ from one another. Notably, when looking at the within-group patterns of aspect performance, the TD group was more likely to change their response than to ask for help, while the other two groups were more likely to ask for help than change their response, although this was only marginally significant for the DS group.

In a study asking 3-5 year old children with DS and CA- and MA-matched controls to perform possible and impossible shape-posting tasks, none of the CA-matched group asked for help, while similar proportions of the DS and MA groups did

(Pitcairn & Wishart, 1994). In addition, the CA-matched group was able to use an alternative strategy of placing the piece where it should go by lifting the lid rather than posting it, while lower numbers of DS and MA groups used this strategy. This is broadly consistent with the present questionnaire outcomes of both atypical groups being more likely to ask for help than to change their response.

The experiences that might lead to a tendency in the atypical groups to ask for help rather than changing what they are doing when they encounter a problem are worthy of consideration. While there are many factors to take into account when considering home and family life, such as parents' personality and education levels (cf. Neitzel & Stright, 2004) parenting style and behaviours have been found to differ between parents of typical and atypical children (Blacher, Baker, & Kaladjian, 2013; Venuti, de Falco, Esposito, Zaninelli, & Bornstein, 2012). De Falco, Venuti, Esposito, and Bornstein (2011) found that parents of children with DS asked fewer questions and made more direct statements to their children than parents of TD children. Gauvain and Huard (1999) found, in an analysis of archival data, that when parenting style was directive when (typical) children were aged 9 years, they were less likely to initiate planning discussions in adolescence than children with other types of parenting styles (authoritative, permissive, uninvolved). Interestingly, Gilmore, Cuskelly, Jobling, and Hayes (2009) demonstrated that children with DS were more persistent with individual problem solving when their mothers directed their behaviour less and supported their independence more on a shared problem-solving task, while this association did not hold for MA-matched controls. Also, adults with learning disabilities were found to be best able to deal with real-life problem solving when they were least dependent on their families (Levine and Langness, 1985, cited in Szepkouski, Gauvain, & Carberry, 1994).

In addition, children with DS are perceived as immature because of their facial characteristics (Fidler & Hodapp, 1999) and their parents' speech to them is at a higher and more varied pitch than controls (Fidler, 2003). One possibility is that children with neurodevelopmental disorders may become more used to receiving help from parents than TD children. This seems intuitively logical in that children with developmental impairments may require more (and more directed) help from parents. This might render them more likely to ask for help from another than to attempt to solve the problem themselves. Szepkouski et al. (1994) found that children with learning difficulties asked for more help than TD children did on a route planning task around a model supermarket.

Help seeking can be seen as a useful strategy, particularly, perhaps, for an individual with a neurodevelopmental disorder, if they are aware that they will be unable to do a particular task without external help (Szepkouski et al., 1994). It may of course also become a maladaptive behaviour: repeated asking for help without attempting to solve the problem would not be a useful strategy in the long term. Szepkouski et al. (1994) suggested that such children may have fewer opportunities to practise planning skills in their everyday lives because of the expectations of others about their low capabilities. While acknowledging that asking for help is an adaptive short-term approach, they warn against encouraging over-reliance on help seeking in intellectually impaired populations as it may further diminish their perceived capabilities. The complex relationship between help seeking, learning and parenting in typical development is discussed by Puustinen, Lyyra, Metsäpelto, and Pulkkinen (2008).

Finally, response aspects did not interact with Group for question 5 (finding a lost possession), while response aspect and group did interact for the other questions. This suggests that this task may not reflect any atypicality in (responding to) problem-solving performance in WS and DS. In contrast, this problem was the one showing relatively poor performance for the DS group in the approach part of the problem. Given that it is somewhat qualitatively different from the other tasks as discussed above, there may be limited ways of responding to something going wrong with this task, which would therefore homogenise the groups' responses.

4.4.4 Aspects associated with reaching the solution

4.4.4.1 Approach

For the approach part of the question, all aspects were significantly associated with reaching the solution. This could point to the nature of the problem operating as a whole: each aspect being needed in order to facilitate success, and failure on any one aspect potentially being enough to elicit task failure.

However, upon breaking the correlations down by question it became apparent that the Steps/Keeptrack variable was only related to reaching the solution in the atypical groups for routine problems (both questions in DS, and only brushing teeth in WS) and not for any individual question in the TD group. Another explanation might therefore be that routine problems, having more easily defined steps, were more likely to determine success or failure if those steps were performed in an incorrect order: putting on shoes before socks or using the brush without putting toothpaste on it may be more difficult to remedy than looking in the same place twice for something that is lost, or putting things away into a wardrobe in a disorganised way. However, it is also possible that this might stem from the wording of the aspects: the Steps/Keeptrack (and What/Ideas) aspects, while based around the same concepts, had different wording for the routine and novel tasks to make them appropriate for the tasks to which they were referring. Adapting the wording in this way may have affected parents' responses to this aspect differentially. This alternative explanation is difficult to support, however, because the What/Ideas aspect did not produce this pattern of response.

4.4.4.2 Response

Change, Lack perseverance and Lose focus were significantly related to Reaching the solution for all groups, while only the WS group showed a relationship between Becoming emotional and Reaching the solution, and only the DS group showed this link for Asking for help. In the study by Fidler, Philofsky, et al. (2005), introduced in Chapter 1, toddlers with DS were less likely to exhibit nonverbal requests for help in relation to objects (e.g., toys) on the Early Social Communication Scales (Elliott, 2007) than a group of TD MA-matched controls.

The frequency of these requests was also significantly related to more efficient reach strategies on an object retrieval task for the DS group. Thus although the paradigms in Fidler and colleagues' study and the present study differ, there is an association between problem-solving success and help seeking in DS in both studies. Fidler et al. speculate about the direction of causality between poor problem solving and low levels of seeking help. Further research would be needed to elucidate this link.

The positive association between asking for help and reaching the solution for the DS group but not the WS group could lend an explanation to the better scores in the DS group than the WS group. Seeking help might serve to bolster the DS scores compared to the WS group: as asking for help was not related to reaching the solution in WS, this group would not gain the advantage afforded to the DS group, thus scoring more poorly.

As all groups were equally likely to ask for help and the DS group were more likely to reach the solution than the WS group, the implication is that the DS group's help seeking was more effective than that of the WS group. As being more likely to become emotional for the WS group was linked to a lower likelihood of reaching the solution, the effectiveness of asking for help in the WS group could be mediated by their tendency to become emotional, which might prevent them from being receptive to the help that is offered. Indeed, non-clinical anxiety (in the typical population) interferes with goal-directed behaviour and attentional control (e.g., Eysenck, Derakshan, Santos, & Calvo, 2007).

The link between becoming emotional and reaching the solution in WS is important given the high rates of anxiety found in children and adults with WS (Leyfer, Woodruff-Borden, Klein-Tasman, Fricke, & Mervis, 2006; Stinton, Elison, & Howlin, 2010), which in children is higher than in the general population (Leyfer, Woodruff-Borden, & Mervis, 2009). Individuals with WS are also more fearful than other groups with intellectual disabilities (Dykens, 2003). Indeed, Einfeld, Tonge, Turner, Parmenter, and Smith (1999) found that the WS group's scores on behaviour disturbance as measured by the Developmental Behaviour Checklist (DBC: e.g., Einfeld & Tonge, 1995) reflected more behavioural and emotional disturbance than that of a DS group. However on the PSQ, all three scales comprising the BRI were significantly related to reaching the solution in both of the atypical groups. This suggests that while the general control of emotions or self may well be linked to problem-solving success for both groups (in line with Gioia et al. (2000a)'s suggestion of the metacognitive scales building on the abilities measured by the regulation scales), it is avoiding becoming emotional about the specific problem being attempted which proves important for the WS group's success.

Although the fears of children with DS were found to be linked to hyperactivity and impulsivity (Evans, Canavera, Kleinpeter, Maccubbin, & Taga, 2005), a link between the regulation of the self and emotions and EF has been demonstrated in WS. Woodruff-Borden et al. (2010) used the BRIEF with parents of 33 children with WS in a longitudinal study, alongside the Anxiety Disorders Interview Schedule (ADIS-P; Silverman & Albano, 1996), used to diagnose anxiety disorders. Those children without an anxiety disorder obtained better scores on the behavioural regulation index (BRI) than those with a diagnosis, and the scores remained stable over time. Rhodes et al. (2010) found that EF tests of planning and attentional flexibility were related to emotional problems in WS, as indicated by scores on the Strengths and Difficulties Questionnaire (SDQ), and that 54% of the sample were classified as showing abnormal levels of impairments on the Emotional Symptoms scale. The relationship between becoming emotional and reaching the solution in WS is consistent with these findings.

4.4.5 Exploring and accounting for better DS than WS performance

Measures of adaptive functioning, such as the VABS, assess everyday life skills including communication, daily living skills and socialisation. Adaptive functioning is known to be poor relative to CA, but good relative to MA, in children with DS (Pennington et al., 2003) with deficits in Communication compared to Socialisation or Daily Living Skills in a group of children with DS aged between 1 and 11 years (Dykens, Hodapp, & Evans, 2006). In contrast, while children with WS also showed poor adaptive functioning on the VABS for their age (Greer, Brown, Pai, Choudry, & Klein, 1997), in adults with WS, VABS scores were poor even in comparison to their full-scale IQ (and improved with age throughout adulthood) (Howlin et al., 2010). Thus, in the mainly teenage and young adult age range sampled in the current study, better performance in DS than in WS reflects existing patterns in the literature regarding adaptive functioning skills. Furthermore, individuals with DS have previously been found to score more highly than individuals with WS on another measure of everyday functioning: social competence, for example with chores, friends and hobbies (Rosner, Hodapp, Fidler, Sagun, & Dykens, 2004).

Sensory modulation (managing one's sensory inputs effectively) allows an individual to select salient aspects of sensory information in the environment to attend to (John & Mervis, 2010). Sensory processing abnormalities (for example, in auditory filtering) have been found in WS (John & Mervis, 2010) and in DS (e.g., Bruni, Cameron, Dua, & Noy, 2010; Wuang & Su, 2011). In Wuang and colleagues' study, sensory processing was also found to be linked to adaptive functioning as measured using the VABS, as well as visual organisation ability, in a sample of 206 children with DS aged 6-13 years. Adaptive functioning on the SIB-R was also measured alongside sensory processing in WS by John and Mervis (2010), and the children with lesser sensory impairments scored better than those with greater sensory impairments on all of the scales: motor skills, social interaction and communication, personal living skills and community living skills. EF accounted for 46% of the variance between higher and lower sensory-impaired groups (John & Mervis, 2010). They argue that difficulties with EF might underlie impairments in flexibly controlling attentional resources to the most appropriate or salient aspects of the environment. This could serve to disrupt focused, goal-directed behaviour or increase the influence of distractions in the WS group, putting them at a disadvantage throughout everyday activities.

Sensory abnormalities are also found in other disorders (e.g., ADHD, autism, Fragile X syndrome; Cascio, 2010) in which EF deficits are also found, although these links have not directly been examined (John & Mervis, 2010). For example, in one study, adaptive behaviour was related to sensory reactivity, which was more severe in groups of young children with autism and fragile X syndrome than in typical or mixed etiology disordered controls (Rogers, Hepburn, & Wehner, 2003).

Thus, the link between sensory processing and adaptive behaviour is present in both WS and DS and is unlikely to be unique to these disorders. As adaptive functioning relative to MA is stronger in DS than in WS, it seems likely that either another factor mediates the relationship between sensory processing and adaptive functioning, or that DS individuals do better on adaptive functioning and everyday problem solving than WS individuals for another reason.

One candidate for mediation is personality or temperament, which takes us back to the link between reaching the solution and becoming emotional on the PSQ in WS that was not found in DS. John and Mervis (2010) also measured Temperament in WS using the Children's Behaviour Questionnaire (CBQ; Rothbart & Ahadi, 1994). Anger/Frustration, Sadness, Falling Reactivity/Soothability, Inhibitory Control and Attentional Focusing were all scored in a more negative way for children with greater sensory impairments than for those with lower sensory impairments.

Therefore, there appears to be a link in WS between sensory processing and both emotional processing and EF (John & Mervis, 2010), and between emotion and EF (Rhodes et al., 2010). John and Mervis (2010) discuss the way in which sensory modulation difficulties can arise from the interaction between task demands in the environment and internal factors such as sensation, emotion and attention, with reference to the Ecological Model of Sensory Modulation (EMSM; Miller, Reisman, McIntosh, & Simon, 2001) which could lead to 'maladaptive behaviours', including anxiety symptoms, difficulties with attention and negative affect.

In Chapter 2, visuospatial working memory, assessed by the backwards block span task, was related to TOL problem solving in the WS group. Importantly, Shackman et al. (2006) drew a link between anxiety and disrupted visuospatial working memory. As individuals with WS are known to suffer with anxiety (Stinton & Howlin, 2012), and to demonstrate poor spatial STM (Jarrold et al., 1998) future research should investigate potential links in the WS population.

4.4.6 Chapter summary

In summary, WS and DS groups scored more poorly on measures of EF and problem solving in everyday life compared to a TD group who were much younger. WS performance was poorer than DS performance, although the two atypical samples were of a similar age range. The WS group showed an unexpected relationship between RCPM and success on the questionnaires, which suggests there was another mediating factor: parental expectations is one possibility. The questionnaire scores for the WS group were also far from where performance would be expected to be placed on the basis of BPVS ability (even though concrete vocabulary is such a strength in WS). EF (BRIEF) scales were related more strongly to novel than routine problems on the PSQ. The Organization of Materials scale was the strongest marker of syndrome-specificity on the BRIEF, with relatively good DS performance marking this group out. Becoming emotional was uniquely related to reaching the solution on the PSQ in the WS group, indicating potential links between anxiety and problem solving for this group, while asking for help was a useful strategy for success in the DS group. The TD group was more likely to change their response than ask for help, while the opposite occurred for the atypical groups, which may be linked to environmental factors associated with growing up with a neurodevelopmental disorder.

CHAPTER 5: RELATIONSHIPS BETWEEN EXPERIMENTAL AND QUESTIONNAIRE MEASURES

5.1 Introduction

In Chapter 2, data were collected for executive functioning (EF) tasks and problem solving on the Tower of London (TOL) in individuals with Williams syndrome (WS) and Down syndrome (DS) and typically developing (TD) controls. Within nine months of this, data were also collected from parents of individuals from the same populations on questionnaire measures of EF and problem solving (Chapter 4). This provides an opportunity to examine the potential links between the two measures in each group.

The representativeness and generalisability of EF tests have been challenged in recent years (e.g., Burgess et al., 2006; see chapter 1 for a discussion of the ecological validity problem). In addition, previous research has found a general lack of association between performance-based tests and rating scales of EF, including the BRIEF, in children with clinical diagnoses including attention deficit/hyperactivity disorder (ADHD) and Tourette's syndrome (TS) (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007; Mahone et al., 2002). However, previous research has found associations in WS between EF tests and behavioural difficulties in everyday life (Rhodes et al., 2010). We also know from Chapter 4 that BPVS scores were related to both questionnaire scores in the WS group and only marginally to the PSQ in the TD group. RCPM was related negatively with questionnaire scores in the WS group, with no significant relationships apparent in the other groups. In this chapter, the experimental and questionnaire data will be correlated to assess whether performance on empirical tasks was related to what was measured when parents were asked about everyday life.

5.2 Method

The method for data collection was described in Chapter 2 for experimental work, and Chapter 4 for the questionnaire data. The participants are the subset identified in Section 4.2.1.2 who provided data in both studies. There were 25 TD children providing both sets of data, 14 participants with DS and 16 participants with WS.

5.3 Results

Scores on the TOL and questionnaires were compared across the three groups of participants providing both experimental and questionnaire data (this was conducted on different groups of participants than in previous chapters, although the same questionnaire outcome is noted in footnote 4 on page 194).

Across groups, there was a significant main effect of TOL score (Welch's $F(2,25.999) = 9.717$, $p = .001$, partial $\eta^2 = .275$), with the TD group scoring better than the DS group ($p = .028$) and the WS group ($p = .002$) with no difference between the scores for the atypical groups ($p = .565$). For the BRIEF percentage score, the main effect of Group ($F(2,52) = 36.222$, $p < .001$, partial $\eta^2 = .582$) was such that the WS score was poorer than the DS score ($p = .002$) and TD score ($p < .001$) and the DS score was poorer than the TD score ($p = .005$). The same pattern was seen for the PSQ percentage score (Welch's $F(2,24.949) = 20.488$, $p < .001$, partial $\eta^2 = .489$; WS<DS, $p = .007$; WS<TD, $p < .001$; DS<TD, $p = .038$). Because of the group differences, correlations between TOL and questionnaire scores were conducted for each group separately. That is, between the TOL score and the BRIEF and PSQ percentage scores, as well as the three indices from the 3-factor approach to the BRIEF used in Chapter 4 (metacognition; behavioural regulation; emotional regulation). None of the correlations were significant for any of the groups ($p > .05$ for all). That is, TOL scores predicted neither everyday EF skills or problem solving, as assessed by parental questionnaires.

Correlations were also run between each MA/EF measure and the reaching-the-solution score for routine and novel problems from the PSQ, with the aim of determining if there were any experimental measures that predicted novel (and routine) problem-solving success; these are reported in Table 5.1. A set of correlations was also run between outcomes of EF tasks and outcomes of questionnaires, for each group. Pairs or groups of measures from the experimental and questionnaire domains were selected that would be likely to demonstrate an association, if the EF tests and the parental report measures were tapping into the same constructs. Correlation analyses for each of these are reported in Table 5.2 below. As for previous chapters, significant correlations ($p \leq .05$) are reported against a green background, while marginally significant correlations ($.05 < p < .10$) are presented against a yellow background. R values for variables that did not meet the assumption of normality are underlined, and Spearman's nonparametric equivalent correlations reported in parentheses if they produced a different pattern of results.

	Reach the solution: routine			Reach the solution: novel		
	TD	DS	WS	TD	DS	WS
	R (*p*)					
BPVS	.281 (.173)	.167 (.568)	.613 (.012)	.518 (.008)	-.032 (.913)	.210 (.435)
RCPM	.180 (.390)	.077 .794)	-.080 (.767)	.230 (.269)	.361 (.204)	-.511 (.043)
TOL	-.217 (.298)	-.435 (.120)	-.162 (.550)	.096 (.650)	.194 (.506)	.254 (.343)
Planning	.207 (.321)	-.124 (.672)	.129 (.646)	.312 (.129)	.195 (.504)	-.271 (.329)
Shifting	.057 (.796)	.209 (.493)	-.010 (.971)	.281 (.194)	.160 (.603)	.071 (.794)
Inhibition: % Opposite	.178 (.394)	.192 (.510)	-.026 (.924)	.412 (.041) (.280, .175)	.332 (.246)	-.064 (.813)
Inhibition: % Change[5]	.139 (.528)	.220 (.451)	.006 (.982)	.448 (.032)	.203 (.487)	-.050 (.853)
Inhibition: RT Same[6]	-.011 (.960)	.139 (.636)	-.057 (.834)	-.065 (.756)	.418 (.137)	-.048 (.860)
Inhibition: RT Opposite[7]	-.201 (.335)	-.126 (.668)	-.010 (.970)	-.303 (.142)	.055 (.852)	-.418 (.107)

Table continues overleaf

[5] The TD correlations presented here are after exclusion of the two TD participants responding in an unusual way to this task (see Section 2.3.5.1). Before exclusion, the correlation for novel problems was marginally significant (r = .346, *p* = .090), while the correlation for routine problems was also non-significant (r = .108, *p* = .607).
[6] Four outliers (2 TD, 1 DS, 1 WS) were identified for RT Same. The significance of the correlations was unchanged upon excluding them from the analysis.
[7] The significance of the correlations was unchanged upon excluding the RT Opposite outliers from the analysis.

Table 5.1 continued	Reach the solution: routine			Reach the solution: novel		
	TD	DS	WS	TD	DS	WS
	R (*p*)					
Inhibition: RT Change[8]	-.107 (.611)	-.253 (.383)	-.018 (.948)	-.290 (.160)	-.466 (.093)	-.516 (.041)
Forwards block span	.034 (.872)	.073 (.814)	-.223 (.407)	.104 (.619)	.094 (.761)	-.044 (.871)
Backwards block span	-.049 (.824)	-.403 (.219)	-.110 (.709)	-.056 (.799)	-.019 (.955)	-.189 (.517)
Forwards digit span	.100 (.770)	.400 (.176)	.263 (.343)	.125 (.713)	.128 (.678)	.281 (.310)
Backwards digit span	-.207 (.541)	-.236 (.437)	-.170 (.562)	.309 (.354)	-.239 (.432)	-.177 (.545)

Table 5.1: Correlations (r, with *p* value in parentheses) between MA/EF measures and routine and novel reach-the-solution score on the PSQ. Where data for at least one variable were non-normal, the r values are underlined. Spearman's correlations are reported if they are different in significance from Pearson's correlations. Significant correlations are displayed against a green background, with marginally significant correlations against a yellow background.

[8] The significance of the correlations was unchanged upon excluding the RT Change outliers from the analysis.

EF test score	BRIEF measure	N (TD, DS, WS)	TD	DS	WS
			R (p)		
Planning	Plan/Organize	25, 14, 16	-.093 (.659)	.052 (.859)	-.081 (.773)
Shifting	Shift	23, 13, 16	.353 (.099)	.043 (.889)	.274 (.304)
Inh: % Opposite	Inhibit	25, 14, 16	-.181 (.386)	-.491 (.074)	-.220 (.412)
Inh: % Change[9]	Inhibit	25, 14, 16	-.302 (.142)	-.357, (.210)	-.212 (.431)
Inh: RT Opposite[10]	Inhibit	25, 14, 16	.315 (.125)	-.145 (.620)	.103 (.704)
Inh: RT Change[11]	Inhibit	25, 14, 16	.322 (.116)	-.032 (.913)	.190 (.480)
Forwards block span	Working Memory	25, 13, 16	-.385 (.057)	.021 (.946)	-.070 (.796)
Backwards block span	Working Memory	23, 11, 14	-.139 (.526)	.191 (.575)	.140 (.634)
Forwards digit span	Working Memory	11, 13, 15	-.436 (.181)	-.371 (.212)	-.640 (.010)
Backwards digit span	Working Memory	11, 13, 14	-.190 (.576)	.279 (.355)	-.285 (.323)

Table 5.2: Correlations (r, with *p* value in parentheses) between EF and questionnaire measures for TD, DS and WS groups. Where data for at least one variable was non-normal, the r values are underlined. Spearman's correlations are reported if they are different in significance from Pearson's correlations. Negative correlations indicate an association between better performance on both measures. Significant correlations are displayed against a green background, with marginally significant correlations against a yellow background. Inh = Inhibition.

[9] Excluding the two TD participants displaying the most extreme unusual behaviour on this task (see Section 2.3.5.1) did not affect the significance of the correlation.

[10] Four outliers (2 TD, 1 DS, 1 WS), two standard deviations away from the mean, were identified for RT Opposite. The significance of the correlations was unchanged upon excluding them from the analysis.

[11] Three outliers (2 TD, 1 DS), two standard deviations away from the mean, were identified for RT Change. The significance of the correlations was unchanged upon excluding them from the analysis.

5.4 Discussion

The lack of significant correlation between the TOL score and the PSQ percentage scores or likelihood of reaching the solution ratings, in any of the groups, indicates that performance on the TOL does not reflect parent-reported problem-solving performance in real-life situations. The ecological validity of the TOL (and EF tests more generally) has previously been challenged (e.g., Burgess et al., 2006; Goel, Grafman, Tajik, Gana, & Danto, 1997) and this finding adds weight to those concerns, and extends them to DS and WS populations. In the neuropsychological literature, responses to the ecological validity problem of EF tests have taken two general forms: verisimilitude and veridicality (Spooner & Pachana, 2006). Verisimilitude is manifested in the development of EF tasks which mirror the types of situations that are encountered in everyday life, while the alternative approach, veridicality, typically uses rating scales to ask about real-life experiences (see Section 4.1.3). The BRIEF is one such measure, designed to give an ecologically valid assessment of EFs, as provided by the caregiver.

The TOL score was also unrelated to the overall BRIEF percentage score, suggesting a lack of overall correspondence with everyday EFs for this task. This is broadly consistent with Toplak et al. (2008) who did not find significant associations between the Stockings of Cambridge (SOC; a TOL-like task) and parent BRIEF scales (although the Teacher Inhibit scale was significantly related), and with Mahone et al. (2002) who also did not find significant TOL associations with BRIEF scales (although the reported correlations were after Bonferroni correction) in a group combining children with ADHD, TS and controls. Anderson et al. (2002) also did not find associations between TOL scores and any of the BRIEF scales for a group of 189 children consisting of individuals with early-treated phenylketonuria, early-treated hydrocephalus, frontal lesions and controls, although some correlations were seen for other EF tasks, such as a Contingency Naming Test, in which several abilities such as switching and response speed are measured (Anderson et al., 2002).

In the current study, correlation analyses were also conducted between scores on EF tests of planning, shifting, inhibition and working memory and parent-reported BRIEF scale scores purported to measure the same construct, in TD, DS and WS groups. Of these, the only correlation reaching significance was between longer forwards digit spans and lower (i.e., better) scores on the Working Memory BRIEF scale in the WS group. Forwards block span was also marginally related to this scale for the TD group, while none of the memory scores showed an association for the DS group. However, there was a weak relationship in the DS group between higher (i.e., poorer) Inhibit scale scores and poorer accuracy on the opposite condition of the inhibition task. So, working memory abilities for TD and WS groups (in different forms) and inhibition abilities for the DS group (tentatively) were observable both on more traditional EF tests and on a real-life rating scale measure of EF, while other EF tasks (planning and shifting) either demonstrated no relationship with BRIEF scale scores or demonstrated a very weak negative relationship in the case of the TD group. This scarcity of relationship between experimental and real-life measures of purportedly the same constructs flags up concerns about the implications that can be drawn about everyday life, based on experimental measures for individuals with WS and DS.

Although it is not possible to rule out the possibility that the lack of correlation is due to the discrepancy in data collection methods, this outcome is not unique to WS and DS. For adolescents with ADHD and controls combined, Toplak et al. (2008) did not find significant associations between a test of inhibition and the BRIEF parent Inhibit scale, or between a test of shifting and the BRIEF parent Shift scale. The working memory BRIEF scale was related to working memory tests, but also to other EF tests, and some other scales were also related to tasks designed to tap into other aspects of EF. Toplak et al. (2013) has more recently published a review article concerning the relationship between EF tests and behavioural rating scales, concluding that, in general, the association between the two types of measure is low and that the two types of measurement tap into different constructs. Anderson and colleagues (2002) suggested that as there is not a reliable relationship between EF tests and BRIEF scales, the cognitive and behavioural aspects of EF may not be equivalent. This is echoed by Chan, Shum, Toulopoulou, and Chen (2008) who discuss the ecological validity difficulties of EF tests in terms of the 'functionality level' at which measurements are taken, with respect to the classification systems of the World Health Organisation – for example, the International Classification of Functioning, Disability and Health (WHO, 2000). They suggest that while an experimental measure might operate at the impairment level (e.g., detecting difficulty switching rules on the WCST), it may not represent the individual's functioning at the wider levels of disability (e.g., problems completing an everyday task satisfactorily) and handicap (e.g., problems with social functions).

The only components of EF that did relate to better outcomes on corresponding BRIEF scales in any way were tests of working memory and inhibition, with only the former reaching significance, for the WS group. Arguably these are fundamental, lower level measures of EF than more complex constructs such as planning or shifting: for example, inhibition and working memory are thought to underlie shifting abilities in typical children (Davidson et al., 2006; also see Mahone et al., 2002). They are also thought to show improvements early in development, particularly with respect to inhibition (Best & Miller, 2010). They have been argued to be impossible to separate behaviourally, although they have been separated at the neurological level (Best, Miller, & Jones, 2009). Therefore, they might reflect more basic processes that can endure the transition from inside to outside the laboratory. Alternatively, difficulties with inhibition (in DS) or working memory (in WS and TD) might be easier to observe in everyday life than more complex constructs like planning or shifting abilities.

Rhodes et al. (2010) also examined relationships between EF scores and parental questionnaires in a WS group, and, in line with the present findings, reported a significant correlation between working memory scores and everyday functioning. However, they also found several links between other EF test scores and everyday problem behaviours, e.g., shifting and the SOC (TOL-like task), which was not found here when EF test scores were correlated with corresponding BRIEF scales. One reason for the discrepancy in findings is that the questionnaire measured employed by Rhodes and colleagues concerned generally social and conduct problems, while the present study's questionnaire examined everyday EFs. One avenue of future research that would be interesting would be to examine the links between the BRIEF and more social behaviours as measured by, for example, the SDQ.

While we saw above that there was a weak relationship between inhibition scores and the Inhibit BRIEF scale in the DS group, when relating EF test scores to reaching-the-solution scores on the PSQ for novel and routine problems, this task was also the only one to show an (albeit, marginal) relationship in the DS group: smaller increases in response time in the opposite condition (RT Change), that is, being less affected by task condition, was related to a greater likelihood of reaching the solution in novel problems. Indeed, inhibition scores were related to reaching the solution in novel problems for all three groups: for the WS group, this was in the same way as in the DS group but with a stronger association, while for the TD group, a drop in accuracy of a smaller magnitude (making the proportional change between conditions less negative) was related to being more likely to reach the solution. This suggests that for all three groups, being able to resist one response in favour of another response which requires more effort is important for real-life problem solving; perhaps, resisting becoming distracted or giving up, to enable more task focus, is the main facilitator for solving a problem when something goes wrong in everyday life for these populations. Best et al. (2009) suggested that inhibition, emerging early in development as it does, indicates that "children's ability to inhibit their prepotent response, and "stop and think" may be an essential first step" (p. 6) thus allowing other EFs to come into play for problem solving. If this is a fundamental ability, it may also be crucial for DS and WS groups. Indeed, although working memory scores were linked to real-life reports of working memory for the WS and TD groups, they were not linked to reaching the solution on the PSQ for any of the groups, potentially reflecting the importance of inhibition.

Some correlations were also seen between MA measures and questionnaire scores. For WS and TD groups, better BPVS scores were related to a greater likelihood of reaching a solution although, notably, this held for routine problems in the WS group but for novel problems in the TD group. This could possibly reflect the use of verbal strategies in the WS group on routine tasks that they encounter daily, while for the TD group this could reflect general cognitive development, although conversely neither CA (from Chapter 4) or RCPM scores were related to reaching the solution in the TD group. In contrast, poorer RCPM scores were associated with a greater likelihood of reaching the solution for novel problems for the WS group. This is consistent with the findings in Chapter 4 of an association between poorer RCPM scores and better outcomes on the PSQ (and BRIEF). The same interpretation offered in Chapter 4 could also apply here: parents' expectations for individuals with better RCPM scores could be higher, and thus relatively worse reports could be garnered for these individuals, with respect to expectations.

Interestingly, the (arguably) most complex tasks (planning, shifting and TOL) were the ones that did not show links either with the corresponding real-life scales on the BRIEF (with the exception of a very slight trend on the shifting task for the TD group, in the opposite direction than would be expected) or with reaching the solution on the PSQ. It might be expected that better performance on more complex tasks generally in the lab would predict better performance in the complex arena of everyday problem solving. However, it appears that for typical children and for individuals with DS and WS, even complex laboratory tasks that demand the integration of several skills did not approximate the complexities of everyday life. This is understandable when we consider that while laboratory tasks are complex in their cognitive demands, everyday tasks may be complex in different ways: e.g., social demands, multitasking demands, and characteristics related to the ill-defined nature of everyday problems (cf., Anderson, 2003). The inherent differences between the testing situation and the home or school environment (Anderson, 2003) are also very likely to play a role in the lack of associations found. One oft-cited quote from Stuss (1987) is that when the experimenter is "becoming the frontal lobes" (p. 174) of the participant, deficits may not be exhibited that are nonetheless apparent outside a testing situation.

CHAPTER 6: GENERAL DISCUSSION

6.1 Aims and overview

The overall aim of this thesis was to investigate problem-solving abilities in typical and atypical development. In the literature, previous research has revealed that: problem solving is thought to rely on executive functions (EFs) in typical development (TD); EF skills have been investigated in Williams syndrome (WS) and Down syndrome (DS) in isolation, and are reported to be impaired, although the profiles of the two disorders are different; the ecological validity of EF tests has been called into question; and the importance of research addressing real life has been emphasised, while problem solving itself has been divided into well-defined and ill-defined problems. The more specific aims of this thesis were therefore to investigate the ways in which EF skills are combined when solving a complex problem in WS and in DS, and to explore real-life problem solving for these individuals, all in comparison to TD individuals. In practice, these were broken down into several constituent aims: first, to compare both EF performance and detailed measures of problem-solving performance across TD, WS and DS groups; second, to investigate and compare the relative contributions of EF skills to complex problem solving; third, to understand and compare problem solving in the real world in these three groups; fourth, to consider aspects of real-life EFs and how they might relate to real-life problem solving; and fifth, to consider the relationship between traditional lab-based testing and parental report measures of EF, to further our understanding of the way in which EF tests and experiences in the real world stand together or, indeed, stand apart.

In this work, a cross-syndrome comparison approach was employed, as advocated by, e.g., Karmiloff-Smith and Farran (2012a). This has allowed an understanding, that would not be available in a study with only one disordered group and one control group, of what is unique to a particular disorder and what is a consequence of having a learning disability.

6.2 Discussion of findings

6.2.1 Summary of findings by chapter

In Chapter 2, it emerged that the DS group's Tower of London (TOL) performance was at a comparable level to that of much younger TD children matched on nonverbal ability, while the WS group's performance was poorer than the matched TD group's. Each group exhibited unique patterns of both EF task performance and associations between EF and TOL tasks, while each atypical group showed some typical-looking patterns of performance as well. Notably, the reaction time (RT) when matching a picture in the Same condition of the inhibition task (before inhibition was required in the Opposite condition) was the most salient measure that was related to TOL score for the DS group, while planning and visuospatial working memory scores were associated with TOL performance in the WS group. While the relationship between RT to the Same condition and TOL was typical for the DS group, the relationship between planning and TOL performance was atypical for the WS group. Adopting the dual approach of comparing group means and analysing developmental trajectories across three groups facilitated interpretations that would otherwise have been unavailable: for example, the DS group appeared not to rely on planning ability for the TOL, while the WS group may have relied upon it more when planning levels were poor.

In Chapter 3, detailed TOL performance was examined across the groups, in terms of the behaviours produced during problem solving as well as the scores and EF-TOL correlations by problem type (easy, medium and difficult). Each of these methods revealed differences in problem-solving approach across groups. For example, the strength of the correlations between EF tasks and TOL score generally increased with problem difficulty for the TD and WS groups but not for the DS group, and the DS group made more rule violations (RVs) than the WS group (but not the TD group). In combination, the latter two findings prompted an interpretation of a different approach by the participants with DS compared to the other groups: rather than increasingly relying on working memory/planning skills as problems became more challenging, the DS group may have turned to the alternative strategy of attempting to reach the solution by breaking the rules of the task (see discussion below).

In Chapter 4, parental ratings of EF and problem solving were compared for TD, DS and WS groups. Overall, WS parent-reported problem solving was poorer than for the DS group, and both atypical groups' scores were poorer than that of another TD group, who were also much younger. Given that language skills in WS are typically stronger than in DS, this result suggests that language does not provide a robust compensatory route to problem solving. Novel problems were more difficult than routine problems on the PSQ, with more BRIEF (EF) scale scores related to success on novel than routine problems. While all groups asked for help to a comparable level, this was only related to reaching the solution for the DS group, while becoming emotional was only related to reaching the solution for the WS group. The atypical groups were more likely to ask for help than change their response, while the opposite was true for the TD group. The Organization of Materials BRIEF scale, which was scored relatively well among the BRIEF scales for the DS group, differentiated their performance from that of the other two groups. Employing a cross-syndrome approach allowed us to determine that this was a characteristic of the DS group rather than a general characteristic of having intellectual impairments overall.

In Chapter 5, the relationships between EF test scores from the lab-based experimental tasks and real-world PSQ outcomes were assessed, to ascertain the extent to which they were tapping into the same constructs. TOL score was unrelated to reaching the solution on the PSQ for each group; moreover, most experimental measures were unrelated to their counterpart BRIEF scales, at the very least, highlighting the different view offered by experimental versus parental report measures, but perhaps also adding to questions about the ecological validity of traditional EF tests (Burgess et al., 2006), and re-emphasising the benefits of also using real-life EF measures (cf. Burgess et al., 1998). The working memory scores were related most strongly to their questionnaire-based counterpart, although this only reached significance for the forwards digit span task for the WS group: therefore, verbal working memory might reflect real-world working memory capabilities in WS. As inhibition test scores (in some form) were weakly related to solving novel problems on the PSQ for each group, this was a tentative indicator that inhibition was the primary experimental indicator of everyday problem-solving success.

6.2.2 Patterns of findings throughout this thesis

6.2.2.1 Processing Speed

Processing speed and response time was an emerging theme from some chapters: reflected by RT to the Same condition of the inhibition task, it was the only measure to be significantly related to TOL score in the DS group. It seemed to be reflected in the relationship between longer planning times and longer times per move in all three groups in Chapter 3, and was generally longer in DS, with longer move times and RTs in the inhibition task for this group. The possibility raised in Chapter 2 of slower processing in DS underlying poorer verbal rehearsal and verbal working memory (as both RTs were slow, and backwards digit span was related to TOL score in DS) was nevertheless weakened by the finding in Chapter 3 that backwards digit span was related most strongly to scores for easy problems, while RT Same was weakly related in all three problem types. If the two skills were related in a functional way for TOL problem solving one would expect similar patterns of reliance for similar problem types.

In WS, after excluding outliers, speed was inversely related to accuracy on the opposite condition of the inhibition task, and there was a suggestion that taking longer to plan was a useful strategy for this group on the TOL as well. Given that poorer TOL scores stemmed mainly from fewer correct trials in the WS group, it may be that TOL errors stemmed from impulsive responding for this group. Indeed, this is echoed by the trajectory analysis of planning and TOL score which revealed that there was more reliance on planning ability for TOL performance when planning was poor for this group, suggesting that when an individual was poor at planning this also hindered their success on the TOL. Speculatively, poor planning and TOL performance could both result from impulsive responding in this group.

Impulsiveness has been found to be a characteristic of WS groups in previous research (Carney et al., 2013; Menghini et al., 2010), and the WS group was slightly (although not significantly) more affected by the Opposite condition of the inhibition task than the TD group (greater changes in RT), indicating possible weaknesses in inhibition relative to nonverbal ability.

6.2.2.2 Working memory and inhibition

While backwards digit span was weakly related to TOL score for all three groups, and backwards block span was additionally related in the WS group (with all four memory tasks related in the TD group), it seems that working memory in general is a good candidate for a component process for TOL success, in some form, for each group. The working memory BRIEF scale was also related to reaching the solution on the PSQ (although marginally in some cases) for all groups, for both routine and novel problems. In addition, working memory was the strongest candidate for being represented both in and out of the laboratory, so seems to have the strongest ecological validity of all of the tasks. Interestingly, however, the working memory EF tests were not related to PSQ problem solving for any of the groups so, in terms of contributing to problem solving, did not span the divide between well-defined and ill-defined arenas. This could reflect a reduced working memory load for the everyday tasks which have been previously encountered.

Inhibition was also (weakly) related to both TOL scores in the WS and DS groups but not the TD group. This could be due to the limited success of this particular inhibition task to detect inhibition skills in the TD children. As inhibition was the score which was (again, weakly) related to PSQ success, this suggests that inhibition is a fundamental enabling ability, for more EFs to be used (cf. Best et al., 2009). Interestingly, the Inhibit scale in the TD group was one of only two scales that were significantly related to novel problem-solving scores on the PSQ, potentially pointing to its importance for problem solving. In contrast, most of the

BRIEF scales were related to the PSQ reaching-the-solution score for the atypical groups. The TD group's lack of reliance on many EF BRIEF scales for PSQ problem-solving success in Chapter 4 might reflect relative ease with the PSQ tasks: if they are easy, the TD group would only need to be focusing (that is, resisting distractions) for success to occur (hence, the relationship between the PSQ success and the Inhibit scale). Interestingly, Task-Monitor was the other scale significantly related to PSQ success for this group, which would fit with these ideas.

6.2.3 Considerations from a neuroconstructivist perspective

This thesis was written from a neuroconstructivist standpoint: that is, rather than emerging from nature or from nurture, outcomes in an individual will have emerged over time through an interaction between processes at many different levels of description (genes, brain, behaviour and environment; Karmiloff-Smith, 1998; 2009). The same scores could be obtained on a test for different reasons in different groups (Karmiloff-Smith, 1998). This is illustrated by the equivalent TOL performance of the DS and TD groups, followed by findings in the DS group of different TOL behaviours (longer times per move; more verbalisations), less EF reliance, and different patterns of EF reliance with increasing problem difficulty than the TD group: the DS group seemed to engage less with EFs with increasing task difficulty, rather than more as the TD group did. In the TOL there are multiple ways of approaching solving: rather than actively engaging in effortful solving (planning ahead, visualising possible moves) success could have been achieved by taking a step by step approach towards the goal (i.e., a hill-climbing strategy). While it is likely that this strategy was used to some degree by members of all the groups, it seems probable that this occurred to a greater extent in the DS group than in the other groups.

Furthermore, the DS group exhibited more RVs than the WS group. This was interpreted as reflecting either: a different strategy involving choosing to make RVs in order to try to solve the task (potentially, compensating for task difficulty by breaking rules, or reflecting low motivation), in line with previous reports of low motivation during problem solving for individuals with DS (e.g., Fidler, Hepburn, Mankin, & Rogers, 2005); or, difficulty in both solving the task and simultaneously remembering to follow the rules (in line with previous findings of DS difficulties with dual-task processing; Lanfranchi et al., 2012). The rule breaking itself could account for the decreased reliance on EFs with increasing difficulty: if participants repeatedly broke the rules and had to stop for the pieces to be reset, opportunities to engage with the task (and use EFs) may have been limited. Future research would be needed to tease apart these possibilities, for example by testing dual-task processing alongside TOL performance and assessing potential links between them.

6.2.3.1 Developmental trajectories

There is a need to measure outcomes developmentally rather than at single points in time (Karmiloff-Smith, 1998, 2009). Considering outcomes using the developmental trajectory approach (Thomas et al., 2009) has allowed an assessment of the way in which one ability is driven by another, which has provided more insight than group comparisons alone would afford. For example, planning was related to TOL score in an atypical way in the WS group, and this would not have been known had the group means alone been compared.

By taking this developmental stance we can appreciate that some patterns of behaviour could reflect compensations over time for areas of weakness. The better score on the Organization of Materials scale in DS could have reflected a compensatory strategy of keeping things in order over development in DS (see Section 4.4.2 for a discussion). Similarly, verbal mediation in the WS group, as a long-term compensatory strategy to use relatively good verbal skills to bootstrap poor nonverbal skills, could have manifested in their relationship between BPVS and both questionnaire and TOL scores, as well as the marginal association between verbalisation and planning times.

By employing a range of different approaches and measures and allowing them to converge, we were able to see both typical and atypical patterns in individuals with neurodevelopmental disorders that may have otherwise not emerged. For example, in Chapter 3, while the reliance on EFs increased in the TD group for more difficult problems, the DS group did not exhibit this pattern, and indeed for some measures demonstrated decreasing EF reliance with increasing problem difficulty. For the WS group, both of these patterns were apparent for different EF/MA measures, so the way that EFs were drawn upon with increasing problem difficulty looked typical for some measures and atypical for others.

6.2.3.2 Theoretical implications

The finding from Chapter 5 that, broadly, little relation emerged between experimental and everyday measures of EF indicates that experimental and questionnaire EF/problem-solving measures were generally not measuring the same constructs. This, on one hand, adds to the body of literature raising concerns about the ecological validity of EF tests (Burgess et al., 2006). On the other, as the BRIEF was designed as one solution to that ecological validity problem, the scarcity of associations perhaps should not be surprising. Indeed these findings are in line with several others, described by Toplak et al. (2013), leading them to conclude that lab-based and real-life measures of EF are tapping into separate entities. The latter, but not the former, require the solver to use background or context in their approach and are often scaffolded by caregivers, to name just a few of the multitude of aspects that are likely to differ between them. In contrast, as mentioned in Chapter 5, a testing situation removes some of the demands that are present in everyday life (cf. Stuss, 1987).

The outcomes can also be considered in light of the different approaches to cognition reviewed in Chapter 1. Briefly, the symbolic approach assumes that an individual interacts with the world using a set of complex rules which could be extended to encompass everyday life (Vera & Simon, 1993). In contrast the situated action approach maintains that cognition in the real world can only be studied in the real world (Greeno & Moore, 1993; Suchman, 1993). As performance on well-defined lab-based tasks were currently not found to bear much relation to performance on ill-defined everyday tasks, this seems to support the situated action approach, in that it is necessary to study real life in its natural context in order to understand it. However, it could be argued that the information-processing approach could also adequately describe everyday problem solving, if enough were known about performance on the task.

As noted in Chapter 2, RT Same was the only measure that was significantly related to TOL performance in DS group, which seems consistent with Karmiloff-Smith (1998)'s suggestion that DS could be a disorder in which the process of 'progressive modularisation', i.e., the development of increasingly specialised abilities, does not occur over time, so that only a general measure of processing was related to complex task performance. Further research would be required to examine this hypothesis, perhaps by examining the comparative interrelatedness of different cognitive abilities in DS and TD.

6.3 Limitations and related considerations

6.3.1 Experimental work

Choosing specific aspects of ability (e.g., verbal: BPVS, or nonverbal: RCPM) for group matching allows the interpretation of results with respect to one aspect of functioning, against the backdrop of uneven cognitive profiles in WS and DS (Farran & Jarrold, 2003). RCPM was chosen as the matching measure so that TOL performance (also being nonverbal) could predict WS/DS performance for the control group's level of nonverbal ability. Here, however, RCPM and TOL scores were unrelated in the atypical groups, which allows for less confidence when judging the expected TOL scores based on nonverbal ability (see Section 2.4.3 for a discussion).

The decision to use mainly nonverbal tasks could have potentially penalised the WS group, as there is evidence that tasks requiring a verbal response are completed with more success for this group (e.g., Atkinson et al., 2003; although see Carney et al., 2013, who found WS impairments in both modalities for inhibition). However, given the opposing profile in DS (e.g., Klein & Mervis, 1999), choosing predominantly verbal tasks would produce the same concern in the reverse direction. Thus, as the TOL task was nonverbal, the most prudent choice was the use of nonverbal tasks. It was possible to also include verbal measures of working memory to allow a fuller picture in this domain.

Some of the experimental limitations stem from the need to consider the issues relevant to designing EF research that is appropriate for children of a large age range as well as for individuals with intellectual disabilities. In the TOL task, the range of problems selected was appropriate for the participants, as demonstrated by a lack of floor or ceiling effects.

One limitation of the threshold procedure that was employed in order to minimise fatigue was that not all participants completed the same number of trials. If the study were to be repeated, a problem set that was completed by all participants would be utilised to yield a richer data set (e.g., for comparisons across problem type as in Chapter 3).

The TOL practice items were relatively easy (2-move) problems, to both ensure task understanding and facilitate a gradual increase in task difficulty. However, it is likely that completing these relatively easy problems did not produce the type of challenging situation in which rule violations might be attempted. Thus, in line with Berg and Byrd (2002)'s suggestion, future iterations of the TOL could include questions about the rules alongside practice items to ensure that the rules had been absorbed by the participant.

It is necessary, particularly in studies which include several tasks, to minimise testing time so as to maintain motivation. In the planning task, pilot testing indicated that the original design of including two trials per delivery length took too long to administer. Thus, only one trial of each length (i.e., each number of houses to be delivered to) was included in the current study. While the TD group obtained a range of scores indicating the task's sensitivity to individuals of different abilities, unfortunately the variation in the number of trials correct was minimal for the DS group, with many participants only able to complete the trial of the same length as the practice trial. It is possible that they were less able to extend the practice trial strategy to future trials, or alternatively that they failed to attend to the additional house on the three house trial. While each house was set up in full view of the participant, it is possible that future studies in which the number of items to be planned for was emphasised ("look, this time there are three houses") might elicit better performance.

Furthermore, having one trial of each length meant that, on failing an item of one length, an opportunity was not available to attempt a trial of that length again: rather, the next trial was more challenging. Future iterations of this task should potentially include more trials at each level, if time constraints allow. Many participants seemed to fail at 3-move trials, either in the first or second phase. The need to introduce a new set of rules, e.g., for phase 2, and start from the easiest delivery length again (2 houses) meant that it could be argued that the first trial of the second phase was not more challenging than the phase 1 trials. Some participants seemed to be able to 'discover the rule' (anecdotally, some of them articulated this rule) in phase 1 of selecting the colours that are needed in the anticlockwise order around the board, indicating that the discovery or non-discovery of this rule might have affected success.

Phases two and three introduced more demanding requirements in order to enhance sensitivity and avoid ceiling effects in the TD group. Reflecting the task impurity problem (see Section 1.2.3.1), it is likely that the working memory demands of this task also increased with the phases: e.g., in phase two, one needs to remember to choose the items in the reverse order but also to place milk boxes underneath cakes. Of the WS group, 40% of participants successfully completed phase one, compared to 60% of the matched TD group. As the block span tasks were poorer in the WS than TD group, it is possible that working memory difficulties may have hindered the WS group in reaching later phases. This speaks to the potential overlap between the block span and planning tasks discussed in Chapter 2.

Finally, the planning task (unlike the TOL) was designed to require the occurrence of planning before execution (delivery). While changes to the order of pieces was not allowed during delivery, in future a rule allowing only one placement of each piece (i.e., disallowing removing an item once it is placed on the van) would increase planning demands further.

In the shifting task, the use of the touchscreen computer was based on an attempt to maintain participants' motivation. In this aspect, the task was largely successful: the touchscreen computer was often engaging for participants to use, and generally elicited a reliable response. However, the requirement to respond by tapping on the screen, combined with the removal of the negative feedback sound for incorrect items (following piloting, to support motivation), unfortunately meant that it was sometimes difficult for participants to distinguish between an incorrect answer and an occasion when they had tapped the screen quickly and the computer had not registered the response. Sometimes this led them to tap the remaining shape, which was not their original choice. As the experimenter was observing and recording the outcome of each trial, this did not cause difficulties with later analysis, but it did mean that reaction time was not a suitable measure for this task. In addition, participants sometimes found it difficult to remember to respond using both hands (which they were asked to do to make it equally easy to select either side of the screen). In future, this task would need to be administered either in a table-top form or using an alternative response method, e.g., a response box.

Developing a shifting task which would be sensitive to a wide range of ability levels was a challenge. Adaptations of the WCST for young children, such as the DCCS, involve telling participants the rule each time it needs to be followed, while the WCST also requires rule discovery and following without input from the experimenter (just feedback regarding the correctness of each trial). The present task was designed to be maximally sensitive to performance, and thus had an easier phase like the DCCS and more difficult phases requiring rule discovery. However, one challenge in developmental EF research is ensuring that what is being measured is constant throughout the age range (Best & Miller, 2010). Indeed, the DCCS has sometimes been described as an inhibition task (Best & Miller, 2010). Shifting is also thought to involve inhibition and working memory (Davidson et al., 2006) and it is easy to see, once again, how the working memory demands of the task would be likely to increase with each phase.

The aim in developing the inhibition task was to measure the type of inhibition necessary for the TOL task: that is, replacing a prepotent response (e.g., simple matching of pieces to positions) with an alternative one (e.g., making a counterintuitive move). Elements of the design were employed in order to minimise the task impurity problem: different sets of images were presented in the two conditions, administered before and after the planning task. The intention was to reduce shifting (e.g., immediately from matching to non-matching) and working memory demands (e.g., to not match the same picture that previously needed to be matched). However, the task did not prove particularly successful for measuring inhibition in the TD group, and may have been under-challenging. Separating the two conditions in time and using different sets of images may have 'diluted' the inhibitory demands.

Similarly, because the pictures that were pressed down for response to the task were attached to the keyboard, the task was simple in its requirement because it is possible to set a rule for responding, e.g., "press my left hand down when I see the grass" and follow that rule during the task. It may be the case that a more complex inhibition task would have accounted for a greater amount of variance in TOL scores. Future work could increase the demands of the inhibition task by: increasing the number of trials of the Same condition (equal numbers of trials were presented in each condition) and thus adding to the prepotency of the response to be inhibited (cf. Kaller et al., 2008's suggestion); including additional matching and inhibiting blocks; using the same set of images in each condition; and not separating the two conditions in time. These changes are likely to increase inhibitory demands, but also other EF demands.

The backwards block span task was selected to allow equivalent forwards and backwards versions of the same task. It should be acknowledged that the particular backwards spatial span employed, from the WNV battery, was designed for use with children from a minimum of 8 years, while the TD age range was from 4 years and above. Anecdotally, young children sometimes needed reminding to produce the sequence backwards and not forwards. However, all but one of the 4-year-olds were able to reverse a sequence of two blocks on an experimental trial, with these reminders. The remaining participant was unfortunately not reminded to produce the sequence backwards and so their data for this task were excluded.

Presenting the backwards memory tasks immediately after the forwards tasks may also have introduced an element of inhibition, or shifting, which may have accounted for some of the difficulties on backwards tasks. It was necessary to use this order of presentation in the digit span tasks to be in keeping with that of the block span tasks, for which the backwards version follows the forwards version in the WNV manual. Also, efforts were made to keep the administration procedures constant between the verbal and visuospatial working memory tasks, although the remaining difference was that for the verbal tasks participants were told how many items to expect, while they were not in the block tasks (with the exception of when it was necessary to inform participants how to reverse three items). This introduced the possibility that block span tasks were more difficult than digit span tasks for this reason. However, given that the TD group was better at block span than digit span tasks in the backwards conditions, and that the atypical groups' patterns of performance followed expected patterns, these concerns were alleviated.

When comparing clinical and typical groups it is useful to control for motor speed on the TOL task (Berg & Byrd, 2002). Thus, the longer response times seen for the DS group could be due to differences in motor speed because this was not controlled for. Also Berg and Byrd (2002) recommended that planning (first move) time includes the motor movement of the first piece, because planning still happens during this first move. Therefore, by not including the first move in the current measure of planning time, not all of the planning may have been picked up. However, they also note that, for example, clinical groups may have longer movement time of the pieces than TD controls, which would introduce a confound, so this has been partially avoided (although the hand movement to make the first move may still have been longer for the atypical groups). Additionally, in light of suggestions that on-line planning is likely to occur (cf. Goel & Grafman, 1995), strictly splitting solution time into planning and execution time may be less achievable than previously imagined.

6.3.2 Questionnaires

While the way in which parents rate their child's behaviour is interesting in itself, reliance on subjective parent reports is a limitation of questionnaire-based measures in general, when one wishes to compare groups. Parental questionnaires provide a direct and efficient way of collecting substantial amounts of data, particularly in the context of working with rare neurodevelopmental disorders such as WS. They also have the advantage of using the knowledge and experience of someone who knows the child in question well. The trade-off is the unknown reliability of parent reports. Rating scales of EF are thought to be susceptible to difficulties related to differences between observers and contexts (Barkley, 2006, cited in Toplak et al., 2013). In the current study, 4.5% of the sample had an inconsistency score on the BRIEF that was rated as 'questionable' while the others fell into the 'acceptable' bracket, and the inconsistency score across the groups was not found to differ significantly. This, of course, speaks to the internal consistency of each parent's ratings rather than the consistency of ratings between respondents. One way to inform reliability for future studies might be to use teacher and parent ratings of the same child; indeed, the BRIEF comprises both parent and teacher forms.

The crucial step between the child's observable behaviour and the questionnaire data is the parent completing it. Parents with high and low expectations are likely to rate the same behaviour differently. One factor that might mediate expectations is whether the parent has already raised a TD child. Although few participants had no siblings, there were no differences on questionnaire scores between them and their counterparts with siblings, indicating that any differences in parent expectations with respect to siblings were not reflected in the questionnaire scores. The role of parent expectations seems to be demonstrated in the TD group's responses. Questionnaire percentage scores were related to CA in the WS and DS groups (and reaching the solution to novel problems was related to CA in the DS group only), but not in the TD group. Given that we might reasonably expect everyday functioning to improve with CA in a TD group, and given the age improvements documented in the BRIEF manual, perhaps TD performance changes are in line with increasing parent expectations with age, producing a constant relationship between them, and therefore no age-related changes.

The TD group was not matched to the atypical groups in Chapter 4. This is because they were recruited to allow the developmental trajectory approach to be employed in a similar way to that used in Chapter 2, and so that the relationships between experimental data from Chapter 2 and questionnaire data could be assessed. However, it does mean that limited conclusions can be drawn about the WS and DS groups' questionnaire performance with respect to controls. Nonetheless, controls were much younger than the atypical groups, so the poorer performance demonstrates considerable difficulties with everyday life in these groups.

The PSQ is a new and innovative measure designed by the author to assess different aspects of problem solving in everyday life. While it was developed with regard to advice from parents and colleagues (see Section 4.1.5.1), it has not been standardised. Partially with this in mind, the BRIEF, a well standardised and validated measure, was administered alongside the PSQ. Future improvements to the questionnaire could include exclusion of the questions concerning making a sandwich and making a telephone call, as these were responded to the least frequently and clearly were not relevant to all of the participants. In future, as part of questionnaire development, parents could rate how often their child attempts a wide variety of activities, and the most commonly chosen ones could be included in the questionnaire. Parents were informed that some tasks were routine and some were novel, so it is possible that they expected novel tasks to be more difficult, and answered accordingly. Of course, by definition of asking about tasks that many people do in everyday life, there will be some element of practice or repetition involved in the tasks, and tasks can be expected to draw on EFs to a lesser degree with repetition (Rabbitt, 1997). On one hand, this points to the usefulness of the response part of the question, which seems to introduce more novelty to the problem; on the other, it is a useful point to acknowledge, as it leads us to question whether any problem in real life is truly novel. Even in experimental tasks, it seems likely that general strategies such as setting up subgoals or vocalising one's behaviour can be utilised in many tasks, and that characteristics of problems will be shared, perhaps requiring planning steps ahead for success. Any systematic study of this sort will need to ask about oft-encountered tasks, which necessarily invokes this type of risk. Another approach might be to engage in naturalistic observation of individuals going about their daily lives and witness their problem solving first hand, although this method would attract substantial practical and methodological challenges. Alternatively, ecologically valid measures of EF have been developed which could constitute a future avenue of research: see discussion below.

6.4 Overall implications and further work

EF impairments in WS and DS were revealed in this thesis as well as in previous studies. However, EF impairments are by no means unique to WS and DS but, as discussed in Chapter 1, exhibited in a range of syndromes (e.g., Pennington & Ozonoff, 1996). Other work (e.g., Kenworthy, Yerys, Anthony, & Wallace, 2008) has investigated similar phenomena in autism, both in laboratory and everyday settings, and there is the potential to extend the current work to other atypical groups, particularly as syndrome-specific patterns have been revealed here.

Procedural (implicit) learning has been found to be less impaired in a DS group than a WS group (Vicari et al., 2000, 2001), which may have lent the DS group an advantage in the real-world arena. Although this would predict more of a DS advantage for routine than for novel tasks, which was not found, it seems possible that it could also apply to general problem-solving strategies: for example, learning that asking for help tends to increase the chances of success. Asking for help was related to problem-solving success for the DS group, but not the WS group. Conversely, for the WS group, being less likely to become emotional was related to better problem-solving success on the PSQ. Interestingly, Toplak et al. (2013) suggested that, based on evidence from previous interventions, increasing structure in everyday life could be helpful for children with ADHD. Many children with WS are known to meet the criteria for ADHD diagnosis (Leyfer et al., 2006), so could also potentially benefit from an increase in structure. Further work would be required to investigate this: it may be that the unpredictability of everyday life could conceivably add to their anxieties, leading to more cognitive difficulties.

In the WS group, BPVS score was related to better questionnaire outcomes generally, and to reaching the solution on the PSQ for routine, but not novel, problems. The implication for parents seems to be that while verbal abilities may be an indicator of everyday functioning more generally, and may predict their child's success on well-practised tasks, they may not be a helpful indicator of an individual's novel problem-solving abilities. In addition, as the WS group's BPVS scores were better than the RCPM-matched TD group's BPVS scores, who nonetheless scored better on the TOL than the WS group, it is clear that verbal abilities do not index problem solving in a straightforward way.

Some suggestions can be gleaned from the present work in moving towards intervention studies, although these should be considered preliminary given the infancy of the research area. First, given the poor level of agreement between experimental and everyday measures, it does not seem likely that increasing performance on laboratory tasks would translate to improvements on real-world tasks. Thus, basing interventions on real-world tasks seems like a prudent approach (see Semel & Rosner, 2003). Of course, we cannot assume that interventions on specific tasks would be generalisable to other tasks (Karmiloff-

Smith, personal communication, October 2013). Second, having said that, the characteristics of the way that the groups approached problem solving might lend something to future approaches: for example, in knowing that individuals with DS are likely to break the rules when a task is difficult, this could be an important indicator to parents or teachers when a child is experiencing difficulty, losing motivation or is stuck on a problem. While there are not usually explicit 'rules of the game' in real life, off-task behaviours could be an indicator of difficulty. Indeed, individuals with DS have previously been found to use social behaviours as distractors in difficult tasks (Wishart, 1993b).

Third, because becoming emotional was related to problem-solving success in the WS group, further research is warranted in investigating the impact of anxiety levels on problem-solving success for this group. As anxiety was related to BRIEF scale scores (John & Mervis, 2010), it seems plausible that it would also show links with the ability to reach a problem's solution. If so, interventions that target reducing anxiety levels during solving could lead to improved outcomes. Fourth, as performance on the experimental inhibition task was related to everyday problem solving in all groups, becoming distracted could (unsurprisingly) be a basic barrier to success, so interventions that either reduce external distractions or increase an individual's ability to disregard them may prove helpful.

Through carrying out this work, other areas of research have presented themselves that could have had a role to play in the current outcomes and may also be fruitful for further research. One factor is the characteristics of family life and parenting styles: as discussed in Chapter 4, there may be reasons related to family life for the pattern of responses gathered (i.e., the atypical groups being more likely to ask for help than change their response to a problem). It is likely that growing up with a disorder would exert an impact on the environment: for example, parents of children with learning disabilities respond to situations differently (e.g., more cautiously) than parents of a typical child (Karmiloff-Smith, 2009). Indeed, parenting looks different for parents of children with neurodevelopmental disorders (Venuti et al., 2012) and could play a role in the way, for example, that they later seek help, as occurs in the typical population (Gauvain & Huard, 1999). Individuals with WS and DS often experience sleep-related difficulties (Annaz, Hill, Ashworth, Holley, & Karmiloff-Smith, 2011) and the potential impact on the current results is unknown. Additionally, the social profile associated with the two syndromes may have had an impact on the present results and is likely to interact with other factors in a real-world context: for example, as the possibility of a role for low task motivation in DS (e.g., Vlachou & Ferrell, 2000, cited in Fidler, 2005) in both the elevated levels of RVs (compared to the WS group) and poor performance on the 'finding a lost possession' question on the PSQ in the DS group was noted, collecting parent ratings of motivation alongside measures of problem-solving performance may be informative. Sensory processing has been found to be atypical in both WS and DS groups and related to EF difficulties in WS (John & Mervis, 2010), and seems worthy of further investigation with respect to combining EFs to solve problems. Finally, attention is known to be atypical in DS and WS, with 'sticky fixation' (difficulties in disengaging attention, or prolonged gaze) in WS at various stages of life (Cornish, Scerif, & Karmiloff-Smith, 2007; Lense, Key, & Dykens, 2011; Mervis et al., 2003; Riby & Hancock, 2008), while infants with DS have impaired sustained attention in infancy (Brown et al., 2003) but not in childhood (Cornish et al., 2007). Breckenridge, Braddick, Anker, Woodhouse, and Atkinson (2012) also found that attention profiles in WS and DS are different. Of course, attending to the task at hand would be crucial for success, particularly if it is a problem which is demanding, so difficulties with attention are likely to impact problem-solving success. Eye-tracking methods would be an interesting future avenue, and are likely to be a good measure of attention on the TOL task (Berg & Byrd, 2002), perhaps in measuring looks to the goal state, as in Hoffman et al. (2003)'s study using the block design task.

It also seems likely that several of these factors will impact on one another for problem-solving success: for example, anxiety is linked to attention in the typical population (Eysenck et al., 2007), EF is related to sensory processing and to anxiety in WS (John & Mervis, 2010), and while inhibition could be interpreted as the ability to ignore distractions, this could just as readily be described as maintaining attention. Ideally, then, future work should take account of several factors simultaneously. Problem solving is complex and multifaceted (Berg et al., 2010) and should be treated as such.

Future work could also set out to capture wider elements of problem solving. Some studies have examined insight in tower-based tasks (Fireman, 1996; Luwel, Siegler, & Verschaffel, 2008). Ill-defined problems, requiring such additional processes as problem recognition and definition, have also been relatively under-studied (Pretz et al., 2003). This type of approach has been attempted previously with respect to problem solving that is understood as alleviating a social conflict or troubling scenario: for example in the Purdue Elementary Problem-Solving Inventory (Feldhusen, Houtz, & Ringenbach, 1972), children were shown cartoons of problem scenarios and asked to go through a series of stages, e.g., selecting from a multiple choice what the problem was. Similarly, Channon, Charman, Heap, Crawford, and Rios (2001) conducted a study in which people with Asperger's syndrome were shown videos of socially awkward scenarios and asked to answer questions, e.g., about the most appropriate solution.

This type of more social approach to problem solving may prove particularly helpful for populations such as WS, as individuals are known to experience difficulties maintaining social relationships (Stinton & Howlin, 2012). Relatedly, problem solving which is carried out collaboratively probably happens frequently in everyday life, and so future research should extend the current investigations to observing the way that people with neurodevelopmental disorders engage in this type of activity, e.g., in parent-child dyads, as in Fagot and Gauvain (1997)'s study. This would be an appropriate arena to investigate the impact of personality, and parenting style, mentioned above.

As noted earlier, a main finding of this thesis is the general scarcity of association between EF and TOL tests and their everyday life, questionnaire counterparts. This indicates that tests and questionnaires were generally not measuring the same constructs, in line with Toplak et al. (2013)'s assertions, and extends similar findings, for example, in ADHD (Bodnar et al., 2007; Toplak et al., 2008) and children with other conditions including brain damage (Anderson et al., 2002) to individuals with WS and DS.

Some authors have considered ways of measuring performance that is relevant to everyday life in an empirical way. Several tasks have been designed to represent real-life demands more faithfully than traditional tests of EF, taking the verisimilitude approach to the ecological validity problem (see Section 4.1.3). Szepkouski et al. (1994) examined planning performance in learning disabled and TD children on a route-planning task set in a model of a supermarket. Another example is the Behavioural Assessment of the Dysexecutive Syndrome test battery; BADS (Wilson, Alderman, Burgess, Emslie, & Evans, 1996; Wilson, Evans, Emslie, Alderman, & Burgess, 1998). This test includes a variety of tasks designed to measure daily executive functioning to aid rehabilitation in neuropsychological patients with Dysexecutive syndrome (DES). In the test, individuals are asked to judge how long certain everyday activities might take, such as blowing up a balloon or visiting the dentist; to switch between responding to the colour of a playing card to responding to whether it matched the previous one they saw or not; to use several pieces of equipment together to remove a piece of cork from a tube; to plan an efficient search strategy; to plan an efficient route around a zoo; and to schedule their own activities in a simplified version of the Six Elements Test (Shallice & Burgess, 1991a). The BADS was found to possess better ecological validity than more standard EF tests (Norris & Tate, 2000). It also includes the Dysexecutive Questionnaire (DEX; Burgess, Alderman, Wilson, Evans, & Emslie, 1996) which can be used to assess executive difficulties in everyday life, and is related to BADS score in brain-damaged (but not schizophrenic) patients (Evans, Chua, McKenna, & Wilson, 1997), and in the children's version of each (Emslie, Wilson, Burden, Nimmo-Smith, & Wilson, 2003). Thus, it may be fruitful for future work to assess individuals with WS and DS on ecologically valid tasks in order to gain an observational set of data of problem solving that is more closely related to people's real lives.

However, when Willner, Bailey, Parry, and Dymond (2010) used the BADS-C and the Cambridge Executive Functioning Assessment (CEFA) to assess EF in individuals with mild to moderate intellectual impairments, they found that the BADS-C was less appropriate than the CEFA because there were some floor effects associated with it. As the CEFA uses more traditional EF measures, such as the TOL and verbal fluency tests, the suitability of the more ecologically valid tasks for populations with neurodevelopmental disorders is as yet unknown. Although these ecologically valid tasks may measure skills required for everyday living, conducting them requires setting up a testing situation. EF requirements may be reduced when an individual engages in a traditional testing paradigm (Bernstein and Waber, 1990, cited in Gioia, Isquith, Kenworthy, et al., 2002). Tasks that establish verisimilitude in this way may be able to avoid some of the ecological validity concerns associated with traditional tests of EF, but they, and indeed any experimental measure, must by definition provide the participant with a task, some rules or restrictions, and a goal, within a distinct paradigm. It is therefore also important to understand individuals' functioning when they are not given a set of instructions in this way; to gather measurements of actual real-life experiences, e.g., through questionnaire studies.

6.5 Final thoughts

This is the first time that the way in which EFs relate to complex problem solving has been compared in Williams syndrome, Down syndrome and typical development. Differences were revealed, not only in the pattern of EF impairments between TD, DS and WS groups matched for nonverbal ability, but in the extent to which different aspects of EF were relied upon for TOL problem solving and the behavioural characteristics exhibited across the groups. Combining several methods and considering outcomes developmentally has led to interpretations that would not otherwise have emerged. The DS and TD groups may have achieved equivalent TOL scores through different cognitive processes. The DS and WS groups approached the task differently, with greater numbers of RVs in the DS group and less reliance on EFs with increasing task difficulty in this group, potentially indicating lower levels of task engagement with increasing difficulty in the DS group, and failure due to visuospatial working memory difficulties in the WS group. Real-life measures of EF and problem solving also revealed syndrome-specific findings: the BRIEF Organization of Materials scale differentiated DS from WS and TD groups, while the atypical groups' problem-solving success was related to different responses to a problem occurring, with help seeking emerging as important for DS, and avoiding becoming emotional for WS. In line with previous findings, a discrepancy between empirical and everyday measures of EF was revealed: the need to consider the real world when assessing EF has been extended for the first time to WS and DS. The importance of using a range of measures, both experimentally and in relation to real life, to understand performance has been emphasised. Future work should examine collaborative and ecologically valid measures of EF and problem solving for these populations, and explore the potential link between anxiety and problem-solving success in WS.

REFERENCE LIST

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cognitive Science, 26*(1), 39-83. doi: http://dx.doi.org/10.1016/S0364-0213(01)00058-1

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*(4), 355.

Anderson, J. R. (2012). Tracking problem solving by multivariate pattern analysis and Hidden Markov Model algorithms. *Neuropsychologia, 50*(4), 487-498. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2011.07.025

Anderson, J. R., & Lebiere, C. (2003). The Newell Test for a theory of cognition. *Behavioral and Brain Sciences, 26*(05), 587-601. doi: 10.1017/S0140525X0300013X

Anderson, P. (2003). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82.

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London Test: Validation and standardization for pediatric populations. *Clinical Neuropsychologist, 10*(1), 54-65.

Anderson, V., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology, 8*(4), 231-240.

Annaz, D., Hill, C. M., Ashworth, A., Holley, S., & Karmiloff-Smith, A. (2011). Characterisation of sleep problems in children with Williams syndrome. *Research in Developmental Disabilities, 32*(1), 164-169.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia, 44*(12), 2259-2269. doi: 10.1016/j.neuropsychologia.2006.05.010

Atkinson, J., Braddick, O., Anker, S., Curran, W., Andrew, R., Wattam-Bell, J., & Braddick, F. (2003). Neurobiological Models of Visuospatial Cognition in Children With Williams Syndrome: Measures of Dorsal-Stream and Frontal Function. *Developmental Neuropsychology, 23*(1-2), 139-172. doi: 10.1080/87565641.2003.9651890

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 302*(1110), 311-324.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baker, K., Segalowitz, S. J., & Ferlisi, M. C. (2001). The Effect of Differing Scoring Methods for the Tower of London Task on Developmental Patterns of Performance. *The Clinical Neuropsychologist, 15*(3), 309-313. doi: 10.1076/clin.15.3.309.10273

Ball, G., Stokes, P. R., Rhodes, R. A., Bose, S. K., Rezek, I., Wink, A.-M., . . . Turkheimer, F. E. (2011). Executive Functions and Pre-Frontal Cortex: A Matter of Persistence? *Frontiers in Systems Neuroscience, 5*. doi: 10.3389/fnsys.2011.00003

Banich, M. T. (2009). Executive Function: The Search for an Integrated Account. *Current Directions in Psychological Science, 18*(2), 89-94. doi: 10.1111/j.1467-8721.2009.01615.x

Baughman, F., & Cooper, R. (2007). Inhibition and young children's performance on the Tower of London task. *Cognitive Systems Research, 8*(3), 216-226.

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition, 119*(2), 301-306.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). I. The neurocognitive profile of Williams syndrome: a complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience, 12*(Supplement 1), 7-29.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams Syndrome: An Ununusal Neuropsychological Profile. In S. H. Broman & J. Grafman (Eds.), *Atypical Cognitive Deficits in Developmental Disorders: Implications for Brain Function* (pp. 23-56). Hillsdale, NJ: Lawrence Erlbaum Associates.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology, 24*(5), 586-604.

Berg, W. K., Byrd, D. L., McNamara, J. P. H., & Case, K. (2010). Deconstructing the tower: Parameters and predictors of problem difficulty on the Tower of London task. *Brain and Cognition, 72*(3), 472-482.

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development, 81*(6), 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*(3), 180-200.

Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual Differences in Cognitive Planning on the Tower of Hanoi Task: Neuropsychological Maturity or Measurement Error? *Journal of Child Psychology and Psychiatry, 42*(4), 551-556. doi: 10.1111/1469-7610.00749

Bittles, A. H., Bower, C., Hussain, R., & Glasson, E. J. (2007). The four ages of Down syndrome. *The European Journal of Public Health, 17*(2), 221-225. doi: 10.1093/eurpub/ckl103

Blacher, J., Baker, B. L., & Kaladjian, A. (2013). Syndrome specificity and mother-child interactions: examining positive and negative parenting across contexts and time. *J Autism Dev Disord, 43*(4), 761-774. doi: 10.1007/s10803-012-1605-x [doi]

Bodnar, L. E., Prahme, M. C., Cutting, L. E., Denckla, M. B., & Mahone, E. M. (2007). Construct Validity of Parent Ratings of Inhibitory Control. *Child Neuropsychology, 13*(4), 345-362. doi: 10.1080/09297040600899867

Breckenridge, K., Braddick, O., Anker, S., Woodhouse, M., & Atkinson, J. (2012). Attention in Williams syndrome and Down's syndrome: performance on the new early childhood attention battery. *British Journal of Developmental Psychology, 31*(2), 257-269. doi: 10.1111/bjdp.12003

BRIEF (Behavior Rating Inventory of Executive Function). (2012). Retrieved 25.06.13, 2013, from http://www4.parinc.com/Products/Product.aspx?ProductID=BRIEF

Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and Psychopathology, 19*(01), 97-127.

Brock, J., Jarrold, C., Farran, E. K., Laws, G., & Riby, D. M. (2007). Do children with Williams syndrome really have good vocabulary knowledge? Methods for comparing cognitive and linguistic abilities in developmental disorders. *Clin Linguist Phon, 21*(9), 673-688. doi: 781355052 [pii]10.1080/02699200701541433 [doi]

Brophy, M., Taylor, E., & Hughes, C. (2002). To go or not to go: inhibitory control in 'hard to manage' children. *Infant & Child Development, 11*(2), 125-140.

Brown, J. H., Johnson, M. H., Paterson, S. J., Gilmore, R., Longhi, E., & Karmiloff-Smith, A. (2003). Spatial representation and attention in toddlers with Williams syndrome and Down syndrome. *Neuropsychologia, 41*(8), 1037-1046.

Bruininks, R. K., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1996). *Scales of Independent Behavior-Revised (SIB-R)*. Itasca, IL: Riverside.

Bruni, M., Cameron, D., Dua, S., & Noy, S. (2010). Reported sensory processing of children with Down syndrome. *Physical & Occupational Therapy in Pediatrics, 30*(4), 280-293. doi: 10.3109/01942638.2010.486962 [doi]

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry, 45*(4), 743-754. doi: 10.1111/j.1469-7610.2004.00268.x

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*(6), 547-558.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M., Dawson, D. R., . . . Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*(2), 194-209.

Burgess, P. W., Alderman, N., Wilson, B. A., Evans, J. J., & Emslie, H. (1996). The Dysexecutive Questionnaire. In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie & J. J. Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome*. Bury St. Edmunds, U.K.: Thames Valley Test Company.

Bush, A., & Beail, N. (2004). Risk factors for dementia in people with Down syndrome: issues in assessment and diagnosis. *American Journal on Mental Retardation, 109*(2), 83-97.

Campos, R., & Sotillo, M. (2012). Mental State Understanding and Social Interaction. In E. K. Farran & A. Karmiloff-Smith (Eds.), *Neurodevelopmental Disorders Across the Lifespan: A Neuroconstructivist Approach*. Oxford: Oxford University Press.

Capone, G., Goyal, P., Ares, W., & Lannigan, E. (2006). *Neurobehavioral disorders in children, adolescents, and young adults with Down syndrome.* Paper presented at the American Journal of Medical Genetics Part C: Seminars in Medical Genetics.

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia, 38*(5), 655-665. doi: http://dx.doi.org/10.1016/S0028-3932(99)00102-5

Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development, 72*(4), 1032-1053. doi: 10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299-319.

Carney, D. P. J., Brown, J. H., & Henry, L. A. (2013). Executive function in Williams and Down syndromes. *Research in Developmental Disabilities, 34*(1), 46-55.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404.

Carr, J. (2008). The Everyday Life of Adults with Down Syndrome. *Journal of Applied Research in Intellectual Disabilities, 21*(5), 389-397. doi: 10.1111/j.1468-3148.2007.00418.x

Carr, J. (2012). Six Weeks to 45 Years: A Longitudinal Study of a Population with Down Syndrome. *Journal of Applied Research in Intellectual Disabilities, 25*(5), 414-422. doi: 10.1111/j.1468-3148.2011.00676.x

Cascio, C. J. (2010). Somatosensory processing in neurodevelopmental disorders. *Journal of neurodevelopmental disorders, 2*(2), 62-69.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23*(2), 201-216. doi: 10.1016/j.acn.2007.08.010 [doi]

Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-Life-Type Problem-Solving in Asperger's Syndrome. *Journal of Autism and Developmental Disorders, 31*(5), 461-469. doi: 10.1023/a:1012212824307

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*(3), 217-227.

Chen, Z., Sanchez, R. P., & Campbell, T. (1997). From beyond to within their grasp: The rudiments of analogical problem solving in 10-and 13-month-olds. *Developmental Psychology, 33*(5), 790.

Chiang, M.-C., Reiss, A. L., Lee, A. D., Bellugi, U., Galaburda, A. M., Korenberg, J. R., . . . Thompson, P. M. (2007). 3D pattern of brain abnormalities in Williams syndrome visualized using tensor-based morphometry. *NeuroImage, 36*(4), 1096-1109.

Clayton, N. (2007). Animal cognition: crows spontaneously solve a metatool task. *Current Biology, 17*(20), R894-895. doi: 10.1016/j.cub.2007.08.028

Cody, H., & Kamphaus, R. (1999). Down Syndrome. In S. Goldstein & C. Reynolds (Eds.), *Handbook of Neurodevelopmental and Genetic Disorders in Children* (pp. 385-405). New York: The Guilford Press.

Cohen, J. (1988) Statistical power for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, R. (2002). Order and disorder in everyday action: the roles of contention scheduling and supervisory attention. *Neurocase, 8*(1-2), 61-79. doi: 10.1093/neucas/8.1.61

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17*(4), 297-338. doi: 10.1080/026432900380427

Cornish, K., Scerif, G., & Karmiloff-Smith, A. (2007). Tracing Syndrome-Specific Trajectories of Attention Across the Lifespan. *Cortex, 43*(6), 672-685.

Costanzo, F., Varuzza, C., Menghini, D., Addona, F., Gianesini, T., & Vicari, S. (2013). Executive functions in intellectual disabilities: A comparison between Williams syndrome and Down syndrome. *Research in Developmental Disabilities, 34*, 1770-1780.

Crnic, L. S., & Pennington, B. F. (2000). Down syndrome: neuropsychology and animal models. In C. Rovee-Collier, L. P. Lipsitt & H. Hayne (Eds.), Progress in Infancy Research volume 1 (Vol. 1, pp. 69-111). Taylor & Francis e-Library, 2010: Lawrence Earlbaum Associates Inc.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of LondonDX: A Standardized Approach to Assessing Executive Functioning in Children. *Archives of Clinical Neuropsychology, 13*(3), 285-301.

Danielsson, H., Henry, L., Rönnberg, J., & Nilsson, L.-G. (2010). Executive functions in individuals with intellectual disability. *Research in Developmental Disabilities, 31*(6), 1299-1304.

Davidson, J. E., & Sternberg, R. J. (2003). *The psychology of problem solving*. Cambridge: Cambridge University Press.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. doi: S0028-3932(06)00055-8 [pii]10.1016/j.neuropsychologia.2006.02.006 [doi]

Davies, M., Howlin, P., & Udwin, O. (1997). Independence and adaptive behavior in adults with Williams syndrome. *American Journal of Medical Genetics, 70*(2), 188-195.

Davies, S. R., Field, A. R., Andersen, T., & Pestell, C. (2011). The ecological validity of the Rey–Osterrieth Complex Figure: Predicting everyday problems in children with neuropsychological disorders. *Journal of Clinical and Experimental Neuropsychology, 33*(7), 820-831.

de Falco, S., Venuti, P., Esposito, G., & Bornstein, M. H. (2011). Maternal and paternal pragmatic speech directed to young children with Down syndrome and typical development. *Infant Behavior and Development, 34*(1), 161-169. doi: http://dx.doi.org/10.1016/j.infbeh.2010.12.002

Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics, 97*(2), 164-171.

Down, J. L. H. (1866). Observations on an ethnic classification of idiots. *London Hospital Reports, 3*, 259-262.

Dunbar, K. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), A companion to cognitive science (pp. 289-298). Oxford, UK: Blackwell Publishing Ltd.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*(3), 257-303.

Dunn, L. M., Dunn, D. M., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., & Burge, B. (2009). *The British Picture Vocabulary Scale* (Third ed.). London: GL Assessment Limited.

Dykens, E., Hodapp, R., & Evans, D. (2006). Profiles and development of adaptive behavior in children with Down syndrome. *Down Syndrome Research and Practice, 9*(3), 45-50.

Dykens, E. M. (2003). Anxiety, fears, and phobias in persons with Williams syndrome. *Developmental Neuropsychology, 23*(1-2), 291-316.

Eckert, M. A., Tenforde, A., Galaburda, A. M., Bellugi, U., Korenberg, J. R., Mills, D., & Reiss, A. L. (2006). To modulate or not to modulate: Differing results in uniquely shaped Williams syndrome brains. *NeuroImage, 32*(3), 1001-1007.

Einfeld, S., Tonge, B., Turner, G., Parmenter, T., & Smith, A. (1999). Longitudinal course of behavioural and emotional problems of young persons with Prader-Willi, Fragile X, Williams and Down syndromes. *Journal of Intellectual and Developmental Disability, 24*(4), 349-354. doi: doi:10.1080/13668259900034111

Einfeld, S. L., & Tonge, B. J. (1995). The Developmental Behavior Checklist: The development and validation of an instrument to assess behavioral and emotional disturbance in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders, 25*(2), 81-104.

Elison, S., Stinton, C., & Howlin, P. (2010). Health and social outcomes in adults with Williams syndrome: Findings from cross-sectional and longitudinal cohorts. *Research in Developmental Disabilities, 31*(2), 587-599.

Elliott, C. D. (1990). *Differential Abilities Scales*. San Diego, CA: Harcourt Brace Jovanovich.

Elliott, C. D. (2007). *Differential Abilities Scales Second Edition*. San Diego, CA: Harcourt Brace Jovanovich.

Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C)*. London, UK: Harcourt Assessment/The Psychological Corporation.

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology, 13*(4), 495-499.

Evans, D. W., Canavera, K., Kleinpeter, F. L., Maccubbin, E., & Taga, K. (2005). The Fears, Phobias and Anxieties of Children with Autism Spectrum Disorders and Down Syndrome: Comparisons with Developmentally and Chronologically Age Matched Children. *Child Psychiatry and Human Development, 36*(1), 3-26. doi: 10.1007/s10578-004-3619-x

Evans, J., Chua, S., McKenna, P., & Wilson, B. (1997). Assessment of the Dysexecutive Syndrome in schizophrenia. *Psychological Medicine, 27*(3), 635-646.

Ewart, A. K., Morris, C. A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., . . . Keating, M. T. (1993). Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nature genetics, 5*(1), 11-16.

Eysenck, M., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*(2), 336.

Eysenck, M., & Keane, M. (2000). Problem Solving: Puzzles, Insight and Expertise *Cognitive Psychology: A Student's Handbook*. Hove: Psychology Press Ltd.

Facon, B., Magis, D., Nuchadee, M.-L., & De Boeck, P. (2011). Do Raven's Colored Progressive Matrices function in the same way in typical and clinical populations? Insights from the intellectual disability field. *Intelligence, 39*(5), 281-291. doi: http://dx.doi.org/10.1016/j.intell.2011.04.002

Facon, B., & Nuchadee, M.-L. (2010). An item analysis of Raven's Colored Progressive Matrices among participants with Down syndrome. *Research in Developmental Disabilities, 31*(1), 243-249.

Fagot, B. I., & Gauvain, M. (1997). Mother-child problem solving: continuity through the early childhood years. *Developmental Psychology, 33*(3), 480-488.

Fahim, C., Yoon, U., Nashaat, N., Khalil, A., El-Belbesy, M., Mancini-Marie, A., . . . Meguid, N. (2012). Williams syndrome: a relationship between genetics, brain morphology and behaviour. *Journal of Intellectual Disability Research, 56*(9), 879-894. doi: 10.1111/j.1365-2788.2011.01490.x

Farran, E. K., & Jarrold, C. (2003). Visuospatial Cognition in Williams Syndrome: Reviewing and Accounting for the Strengths and Weaknesses in Performance. *Developmental Neuropsychology, 23*(1), 173 - 200.

Farran, E. K., Jarrold, C., & Gathercole, S. E. (2001). Block Design Performance in the Williams Syndrome Phenotype: A Problem with Mental Imagery? *Journal of Child Psychology and Psychiatry, 42*(6), 719-728. doi: 10.1111/1469-7610.00768

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Feldhusen, J. F., Houtz, J. C., & Ringenbach, S. (1972). The Purdue Elementary Problem-solving Inventory. *Psychological Reports, 31*(891-901).

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: relations with task difficulty and task performance. *Cognitive Development, 20*(1), 103-120. doi: 10.1016/j.cogdev.2004.11.002

Fidler, D. (2003). Parental Vocalizations and Perceived Immaturity in Down Syndrome. *American Journal on Mental Retardation, 108*(6), 425-434. doi: 10.1352/0895-8017(2003)108<425:pvapii>2.0.co;2

Fidler, D. (2005). The emerging Down syndrome behavioral phenotype in early childhood: Implications for practice. *Infants & Young Children, 18*(2), 86-103.

Fidler, D., Hepburn, S. L., Mankin, G., & Rogers, S. (2005). Praxis skills in young children with Down syndrome, other developmental disabilities, and typically developing children. *The American journal of occupational therapy, 59*(2), 129-138.

Fidler, D., & Hodapp, R. (1999). Craniofacial Maturity and Perceived Personality in Children With Down Syndrome. *American Journal on Mental Retardation, 104*(5), 410-421. doi: 10.1352/0895-8017(1999)104<0410:cmappi>2.0.co;2

Fidler, D., & Nadel, L. (2007). Education and children with Down syndrome: neuroscience, development, and intervention. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(3), 262-271. doi: 10.1002/mrdd.20166 [doi]

Fidler, D., Philofsky, A., Hepburn, S., & Rogers, S. (2005). Nonverbal requesting and problem-solving by toddlers with down syndrome. *American Journal on Mental Retardation, 110*(4), 312-322.

Fireman, G. (1996). Developing a plan for solving problem: A representational shift. *Cognitive Development, 11*(1), 107-122.

Fodor, J. (Ed.). (1983). *The Modularity of mind [electronic resource]: an essay on faculty psychology*: The MIT Press.

Frisch, S., Förstl, S., Legler, A., Schöpe, S., & Goebel, H. (2012). The interleaving of actions in everyday life multitasking demands. *Journal of neuropsychology, 6*(2), 257-269.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*(1–3), 1-34. doi: http://dx.doi.org/10.1016/S0301-0511(00)00051-X

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483-527. doi: http://dx.doi.org/10.1016/0885-2014(95)90024-1

Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*(6), 817-831. doi: 10.1207/s15516709cog2606_5

Gaser, C., Luders, E., Thompson, P. M., Lee, A. D., Dutton, R. A., Geaga, J. A., . . . Korenberg, J. R. (2006). Increased local gyrification mapped in Williams syndrome. *NeuroImage, 33*(1), 46-54.

Gauvain, M., & Huard, R. D. (1999). Family interaction, parenting style, and the development of planning: A longitudinal analysis using archival data. *Journal of Family Psychology, 13*(1), 75-92.

Gibson, D. (1978). *Down syndrome: The psychology of mongolism*. London: Cambridge University Press.

Gilmore, L., Cuskelly, M., Jobling, A., & Hayes, A. (2009). Maternal support for autonomy: Relationships with persistence for children with Down syndrome and typically developing children. *Research in Developmental Disabilities, 30*(5), 1023-1033.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000a). *Behavior Rating Inventory of Executive Function*. USA: Psychological Assessment Resources, Inc.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000b). TEST REVIEW Behavior Rating Inventory of Executive Function. *Child Neuropsychology, 6*(3), 235-238.

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of Everyday Executive Function in Acquired and Developmental Disorders. *Child Neuropsychology, 8*(2), 121-137. doi: 10.1076/chin.8.2.121.8727

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory Factor Analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a Clinical Sample. *Child Neuropsychology, 8*(4), 249-257. doi: 10.1076/chin.8.4.249.13513

Glenn, S., & Cunningham, C. (2005). Performance of young people with Down syndrome on the Leiter-R and British picture vocabulary scales. *Journal of Intellectual Disability Research, 49*(4), 239-244. doi: 10.1111/j.1365-2788.2005.00643.x

Godefroy, O., Cabaret, M., Petit-Chenal, V., Pruvo, J.-P., & Rousseaux, M. (1999). Control Functions of the Frontal Lobes. Modularity of the Central-Supervisory System? *Cortex, 35*(1), 1-20. doi: http://dx.doi.org/10.1016/S0010-9452(08)70782-2

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia, 33*(5), 623-642.

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain, 120*(10), 1805-1822. doi: 10.1093/brain/120.10.1805

Grace-Martin, K. (2008). Can Likert Scale Data ever be continuous? Retrieved 07.05.13, 2013, from http://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J Exp Psychol, 38*(4), 404-411.

Grant, J., Karmiloff-Smith, A., Gathercole, S. A., Paterson, S., Howlin, P., Davies, M., & Udwin, O. (1997). Phonological Short-term Memory and its Relationship to Language in Williams Syndrome. *Cognitive Neuropsychiatry, 2*(2), 81-99. doi: 10.1080/135468097396342

Gray, V., Karmiloff-Smith, A., Funnell, E., & Tassabehji, M. (2006). In-depth analysis of spatial cognition in Williams syndrome: A critical assessment of the role of the LIMK1 gene. *Neuropsychologia, 44*(5), 679-685.

Greeno, J. G., & Moore, J. L. (1993). Situativity and symbols: Response to Vera and Simon. *Cognitive Science, 17*(1), 49-59.

Greer, M. K., Brown, F. R., Pai, G. S., Choudry, S. H., & Klein, A. J. (1997). Cognitive, adaptive, and behavioral characteristics of Williams syndrome. *American Journal of Medical Genetics, 74*(5), 521-525. doi: 10.1002/(sici)1096-8628(19970919)74:5<521::aid-ajmg13>3.0.co;2-e

Groth-Marnat, G., & Teal, M. (2000). Block design as a measure of everyday spatial ability: A study of ecological validity. *Perceptual and Motor Skills, 90*(2), 522-526.

Gunn, D. M., & Jarrold, C. (2004). Raven's matrices performance in Down syndrome: Evidence of unusual errors. *Research in Developmental Disabilities, 25*(5), 443-457. doi: http://dx.doi.org/10.1016/j.ridd.2003.07.004

Hocking, D. R., Bradshaw, J. L., & Rinehart, N. J. (2008). Fronto-parietal and cerebellar contributions to motor dysfunction in Williams syndrome: A review and future directions. *Neuroscience & Biobehavioral Reviews, 32*(3), 497-507.

Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial Breakdown in Spatial Construction: Evidence from Eye Fixations in Children with Williams Syndrome. *Cognitive Psychology, 46*(3), 260-301.

Howlin, P., Davies, M., & Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry, 39*(2), 183-189.

Howlin, P., Elison, S., Udwin, O., & Stinton, C. (2010). Cognitive, Linguistic and Adaptive Functioning in Williams Syndrome: Trajectories from Early to Middle Adulthood. *Journal of Applied Research in Intellectual Disabilities, 23*(4), 322-336. doi: 10.1111/j.1468-3148.2009.00536.x

Hughes, C. (2002). Executive functions and development: emerging themes. *Infant and Child Development, 11*(2), 201-209. doi: 10.1002/icd.297

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036.

Jackowski, A. P., Rando, K., Maria de Araújo, C., Del Cole, C. G., Silva, I., & Tavares de Lacerda, A. L. (2009). Brain abnormalities in Williams syndrome: A review of structural and functional magnetic resonance imaging findings. *European Journal of Paediatric Neurology, 13*(4), 305-316.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The Development of Children's Rule Use on the Balance Scale Task. *Journal of Experimental Child Psychology, 81*(4), 383-416. doi: http://dx.doi.org/10.1006/jecp.2002.2664

Jarrold, C., & Baddeley, A. D. (1997). Short-term memory for verbal and visuospatial information in Down's syndrome. *Cognitive Neuropsychiatry, 2*(2), 101-122.

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1998). Verbal and Nonverbal Abilities in the Williams Syndrome Phenotype: Evidence for Diverging Developmental Trajectories. *Journal of Child Psychology and Psychiatry, 39*(4), 511-523. doi: 10.1111/1469-7610.00347

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1999). Genetically dissociated components of working memory: evidence from Downs and Williams syndrome. *Neuropsychologia, 37*(6), 637-651.

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (2000). Verbal short-term memory deficits in Down syndrome: a consequence of problems in rehearsal? *J Child Psychol Psychiatry, 41*(2), 233-244.

Jarrold, C., Baddeley, A. D., Hewes, A. K., & Phillips, C. (2001). A Longitudinal Assessment of Diverging Verbal and Non-Verbal Abilities in the Williams Syndrome Phenotype. *Cortex, 37*(3), 423-431.

Jawaid, A., Riby, D., Owens, J., White, S., Tarar, T., & Schulz, P. (2011). 'Too withdrawn'or 'too friendly': considering social vulnerability in two neuro-developmental disorders. *Journal of Intellectual Disability Research, 56*(4), 335-350.

Jernigan, T. L., Bellugi, U., Sowell, E., Doherty, S., & Hesselink, J. R. (1993). Cerebral morphologic distinctions between Williams and Down syndromes. *Archives of Neurology, 50*(2), 186.

John, A. E., & Mervis, C. B. (2010). Sensory modulation impairments in children with Williams syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 154C*(2), 266-276.

Jones, K. L., & Smith, D. W. (1975). The Williams elfin facies syndrome: A new perspective. *The Journal of Pediatrics, 86*(5), 718-723.

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., & Adolphs, R. (2000). II. Hypersociability in Williams syndrome. *Journal of Cognitive Neuroscience, 12*(Supplement 1), 30-46.

Jurado, M., & Rosselli, M. (2007). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychology Review, 17*(3), 213-233. doi: 10.1007/s11065-007-9040-z

Kahney, H. (1986a). Getting started on problem solving *Problem solving: A cognitive approach* (pp. 18-23). Milton Keynes, UK: Open University Press

Kahney, H. (1986b). *Problem Solving: A cognitive approach*. Milton Keynes, Philadelphia: Open University Press.

Kaller, C. P., Rahm, B., Bolkenius, K., & Unterrainer, J. M. (2009). Eye movements and visuospatial problem solving: Identifying separable phases of complex cognition. *Psychophysiology, 46*(4), 818-830.

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research, 216*(1), 1-8.

Kaller, C. P., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain and Cognition, 67*(3), 360-370.

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: insights from the Tower of London task. *Cognitive Brain Research, 20*(3), 462-472. doi: 10.1016/j.cogbrainres.2004.04.002

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology, 13*(4), 513-524.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2*(10), 389-398.

Karmiloff-Smith, A. (2006). The tortuous route from genes to behavior: A neuroconstructivist approach. *Cognitive, Affective, & Behavioral Neuroscience, 6*(1), 9-17.

Karmiloff-Smith, A. (2009). Nativism Versus Neuroconstructivism: Rethinking the Study of Developmental Disorders. *Developmental Psychology, 45*(1), 56-63.

Karmiloff-Smith, A., Broadbent, H., Farran, E. K., Longhi, E., D'Souza, D., Metcalfe, K., . . . Sansbury, F. (2012). Social Cognition in Williams Syndrome: Genotype/phenotype Insights from Partial Deletion Patients. *Frontiers in Psychology, 3*. doi: 10.3389/fpsyg.2012.00168

Karmiloff-Smith, A., & Farran, E. (2012a). Future theoretical and empirical directions within a neuroconstructivist framework. In E. Farran & A. Karmiloff-Smith (Eds.), *Neurodevelopmental Disorders Across the Lifespan: A Neuroconstructivist Approach* (pp. 363-371). Oxford: Oxford University Press.

Karmiloff-Smith, A., & Farran, E. (2012b). Introduction Williams syndrome: a model for the neuroconstructivist approach. In A. Karmiloff-Smith & E. K. Farran (Eds.), *Neurodevelopmental disorders across the lifespan: a neuroconstructivist approach*. Oxford: Oxford University Press.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double Dissociations in developmental disorders? Theoretically misconceived, empirically dubious. Discussion, . *Cortex, 39*, 161-163.

Karmiloff-Smith, A., Thomas, M. S. C., Annaz, D., Humphreys, K., Ewing, S., Brace, N., . . . Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: the importance of building developmental trajectories. *J Child Psychol Psychiatry, 45*(7), 1258-1274. doi: 10.1111/j.1469-7610.2004.00322.x [doi]JCPP322 [pii]

Kates, W. R., Folley, B. S., Lanham, D. C., Capone, G. T., & Kaufmann, W. E. (2002). Cerebral growth in Fragile X syndrome: review and comparison with Down syndrome. *Microscopy research and technique, 57*(3), 159-167.

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review, 18*(4), 320-338.

Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological Validity of Neuropsychological Tests: Focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology, 13*(6), 523-534. doi: http://dx.doi.org/10.1016/S0887-6177(97)00038-3

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). ARTICLE WITH PEER COMMENTARIES AND RESPONSE Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449-476.

Kittler, P. M., Krinsky-McHale, S. J., Devenny, D. A., & Conners, F. (2008). Dual-Task Processing as a Measure of Executive Function: A Comparison Between Adults With Williams and Down Syndromes. *American Journal on Mental Retardation, 113*(2), 117-132. doi: 10.1352/0895-8017(2008)113[117:dpaamo]2.0.co;2

Klahr, D. (1985). Solving Problems with Ambiguous Subgoal Ordering: Preschoolers' Performance *Child Development, 56*(4), 940-952.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13*(1), 113-148.

Klein, B. P., & Mervis, C. B. (1999). Contrasting Patterns of Cognitive Abilities of 9- and 10-Year-Olds With Williams Syndrome or Down Syndrome. *Developmental Neuropsychology, 16*(2), 177-196. doi: 10.1207/s15326942dn1602_3

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181.

Koppenol-Gonzalez, G. V., Bouwmeester, S., & Boonstra, A. M. (2010). Understanding Planning Ability Measured by the Tower of London: An Evaluation of Its Internal Structure by Latent Variable Modeling. *Psychological Assessment, 22*(4), 923-934.

Krikorian, R., Bartok, J., & Gay, N. (1994). Tower-of-London procedure - a standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology, 16*(6), 840-850.

Laing, E., & Jarrold, C. (2007). Comprehension of Spatial Language in Williams Syndrome: Evidence for Impaired Spatial Representation of Verbal Descriptions. *Clinical Linguistics and Phonetics, 21*(9), 689-704.

Lanfranchi, S., Baddeley, A., Gathercole, S., & Vianello, R. (2012). Working memory in Down syndrome: is there a dual task deficit? *Journal of Intellectual Disability Research, 56*(2), 157-166. doi: 10.1111/j.1365-2788.2011.01444.x

Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A., & Vianello, R. (2010). Executive function in adolescents with Down Syndrome. *Journal of Intellectual Disability Research, 54*(4), 308-319. doi: 10.1111/j.1365-2788.2010.01262.x

Laws, G. (2002). Working memory in children and adolescents with Down syndrome: evidence from a colour memory experiment. *Journal of Child Psychology and Psychiatry, 43*(3), 353-364. doi: 10.1111/1469-7610.00026

Laws, G., & Lawrence, L. (2001). Spatial representation in the drawings of children with Down's syndrome and its relationship to language and motor development: A preliminary investigation. *British Journal of Developmental Psychology, 19*, 453-473.

Lebiere, C., & Anderson, J. R. (2008). *A connectionist implementation of the ACT-R production system*Paper presented at the Department of Psychology. Paper 33., http://repository.cmu.edu/psychology/33.

Lee, N. R., Fidler, D., Blakeley-Smith, A., Daunhauer, L., Robinson, C., & Hepburn, S. (2011). Caregiver report of executive functioning in a population-based sample of young children with Down syndrome. *American journal on intellectual and developmental disabilities, 116*(4), 290-304.

Leeds, U. o. (n.d.). Guide to the Design of Questionnaires Retrieved 25.06.13, 2013, from http://iss.leeds.ac.uk/info/312/surveys/217/guide_to_the_design_of_questionnaires/3

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80.

Lejeune, J., Gauthier, M., & Turpin, R. (1959). Human chromosomes in tissue cultures. *Comptes rendus hebdomadaires des séances de l'Académie des sciences, 248*(4), 602-603.

Lense, M. D., Key, A. P., & Dykens, E. M. (2011). Attentional disengagement in adults with Williams syndrome. *Brain and Cognition, 77*(2), 201-207.

Leyfer, O., Woodruff-Borden, J., & Mervis, C. B. (2009). Anxiety disorders in children with williams syndrome, their mothers, and their siblings: implications for the etiology of anxiety disorders. *Journal of Neurodevelopmental Disorders, 1*(1), 4-14.

Leyfer, O. T., Woodruff-Borden, J., Klein-Tasman, B. P., Fricke, J. S., & Mervis, C. B. (2006). Prevalence of psychiatric disorders in 4 to 16-year-olds with Williams syndrome. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 141B*(6), 615-622. doi: 10.1002/ajmg.b.30344

Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology, 107*(4), 438-451.

Lincoln, A. J., Courchesne, E., Kilman, B. A., & Galambos, R. (1985). Neuropsychological correlates of information-processing by children with Down syndrome. *American Journal of Mental Deficiency, Vol 89*(4), 403-414.

Lögdbergi, B., & Brun, A. (1993). Prefrontal neocortical disturbances in mental retardation. *Journal of Intellectual Disability Research, 37*(5), 459-468. doi: 10.1111/j.1365-2788.1993.tb00316.x

Lott, I. T., & Head, E. (2001). Down syndrome and Alzheimer's disease: A link between development and aging. *Mental Retardation and Developmental Disabilities Research Reviews, 7*(3), 172-178. doi: 10.1002/mrdd.1025

Loughborough, U. (n.d.). Questionnaire Design Retrieved 25.06.13, 2013, from http://www.lboro.ac.uk/media/wwwlboroacuk/content/library/downloads/advicesheets/questionnaire.pdf

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36*(3), 273-293.

Luwel, K., Siegler, R. S., & Verschaffel, L. (2008). A microgenetic study of insightful problem solving. *Journal of Experimental Child Psychology, 99*(3), 210-232. doi: http://dx.doi.org/10.1016/j.jecp.2007.08.002

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., . . . Denckla, M. B. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology, 17*(7), 643-662. doi: http://dx.doi.org/10.1016/S0887-6177(01)00168-8

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *TEA-Ch: the test of everyday attention for children*: Thames Valley Test Company Bury St. Edmunds, England.

Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Williams Syndrome: A Critical Review of the Cognitive, Behavioral, and Neuroanatomical Phenotype. *Journal of Child Psychology and Psychiatry, 49*(6), 576-608.

Masson, J. D., Dagnan, D., & Evans, J. (2010). Adaptation and validation of the Tower of London test of planning and problem solving in people with intellectual disabilities. *Journal of intellectual disability research : JIDR, 54*(5), 457-467.

Menghini, D., Addona, F., Costanzo, F., & Vicari, S. (2010). Executive Functions in Individuals with Williams Syndrome. *Journal of Intellectual Disability Research, 54*(5), 418-432.

Mervis, C. B., & John, A. E. (2008). Vocabulary abilities of children with Williams syndrome: Strengths, weaknesses, and relation to visuospatial construction ability. *Journal of Speech, Language, and Hearing Research, 51*(4), 967.

Mervis, C. B., & John, A. E. (2010). Cognitive and Behavioral Characteristics of Children With Williams Syndrome: Implications for Intervention Approaches. *American Journal of Medical Genetics Part C-Seminars in Medical Genetics, 154C*(2), 229-248. doi: 10.1002/ajmg.c.30263

Mervis, C. B., Morris, C. A., Klein-Tasman, B. P., Bertrand, J., Kwitny, S., Appelbaum, L. G., & Rice, C. E. (2003). Attentional Characteristics of Infants and Toddlers With Williams Syndrome During Triadic Interactions. *Developmental Neuropsychology, 23*(1-2), 243-268. doi: 10.1080/87565641.2003.9651894

Mervis, C. B., Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P., & Armstrong, S. C. (2000). The Williams Syndrome Cognitive Profile. *Brain and Cognition, 44*(3), 604-628.

Meyer-Lindenberg, A., Hariri, A. R., Munoz, K. E., Mervis, C. B., Mattay, V. S., Morris, C. A., & Berman, K. F. (2005). Neural correlates of genetically abnormal social cognition in Williams syndrome. *Nature Neuroscience, 8*(8), 991-993.

Meyer-Lindenberg, A., Kohn, P., Mervis, C. B., Kippenhan, J. S., Olsen, R. K., Morris, C. A., & Berman, K. F. (2004). Neural Basis of Genetically Determined Visuospatial Construction Deficit in Williams Syndrome. *Neuron, 43*(5), 623-631.

Meyer-Lindenberg, A., Mervis, C. B., Sarpal, D., Koch, P., Steele, S., Kohn, P., . . . Kippenhan, S. (2005). Functional, structural, and metabolic abnormalities of the hippocampal formation in Williams syndrome. *Journal of Clinical Investigation, 115*(7), 1888.

Miller, L., Reisman, J. E., McIntosh, D., & Simon, J. (2001). An ecological model of sensory modulation: Performance of children with fragile X syndrome, autistic disorder, attention-deficit/hyperactivity disorder, and sensory modulation dysfunction. *Understanding the nature of sensory integration with diverse populations*, 57-88.

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*(1), 49-100.

Morris, C. A., & Mervis, C. B. (1999). Williams Syndrome. . In S. Goldstein & C. R. Reynolds (Eds.), *Handbook of Neurodevelopmental and Genetic Disorders in Children* (pp. 555–590). New York: The Guilford Press.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the tower of London test. *Neuropsychologia, 31*(12), 1367-1378.

Multon, K. (2010). Interrater Reliability. In N. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 627-629). Thousand Oaks, CA: SAGE Publications, Inc.

Mushiake, H., Sakamoto, K., Saito, N., Inui, T., Aihara, K., Tanji, J., . . . Sakurada, S. (2009). Chapter 1 Involvement of the Prefrontal Cortex in Problem Solving *International Review of Neurobiology* (Vol. Volume 85, pp. 1-11): Academic Press.

Nadel, L. (2003). Down's syndrome: a genetic disorder in biobehavioral perspective. *Genes, Brain and Behavior, 2*(3), 156-166.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.

Natsopoulo, D., Christou, C., Koutselini, M., Raftopoulos, A., & Karefillidou, C. (2002). Structure and coherence of reasoning ability in Down Syndrome adults and typically developing children. *Research in Developmental Disabilities, 23*(4), 297-307.

Neitzel, C., & Stright, A. D. (2004). Parenting behaviours during child problem solving: The roles of child temperament, mother education and personality, and the problem-solving context. *International Journal of Behavioral Development, 28*(2), 166-179. doi: 10.1080/01650250344000370

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*(4), 313-324.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). *Report on a general problem-solving program.* Paper presented at the IFIP Congress.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Newman, S. D., & Pittman, G. (2007). The Tower of London: A study of the effect of problem structure on planning. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 333-342. doi: 10.1080/13803390701249051

Norman, D. A. (1993). Cognition in the Head and in the World: An Introduction to the Special Issue on Situated Action. *Cognitive Science, 17*(1), 1-6.

Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behaviour. Center for Human Information Processing (Technical Report No. 99) (1980) Reprinted in revised form in. In R. J. Davidson, G. E. Schwartz & D. D. Shapiro (Eds.), *Consciousness and Self-Regulation* (Vol. 4). New York: Plenum Press, 1986.

Norris, G., & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation, 10*(1), 33-45.

Odhuba, R. A., Broek, M., & Johns, L. (2005). Ecological validity of measures of executive functioning. *British Journal of Clinical Psychology, 44*(2), 269-278.

Osório, A., Cruz, R., Sampaio, A., Garayzábal, E., Martínez-Regueiro, R., Gonçalves, Ó. F., . . . Fernández-Prieto, M. (2012). How executive functions are related to intelligence in Williams syndrome. *Research in Developmental Disabilities, 33*(4), 1169-1175.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28*(10), 1021-1034.

Paterson, S. J., Brown, J. H., Gsodl, M. K., Johnson, M. H., & Karmiloff-Smith, A. (1999). Cognitive modularity and genetic disorders. *Science, 286*(5448), 2355-2358. doi: 8098 [pii]

Pennington, B. F., Moon, J., Edgin, J., Stedron, J., & Nadel, L. (2003). The Neuropsychology of Down Syndrome: Evidence for Hippocampal Dysfunction. *Child Development, 74*(1), 75-93.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*(1), 51-87.

Pezzini, G., Vicari, S., Volterra, V., Milani, L., & Ossella, M. T. (1999). Children with williams syndrome: Is there a single neuropsychological profile? *Developmental Neuropsychology, 15*(1), 141-155. doi: 10.1080/87565649909540742

Phillips, C. E., Jarrold, C., Baddeley, A. D., Grant, J., & Karmiloff-Smith, A. (2004). Comprehension of Spatial Language Terms in Williams Syndrome: Evidence for an Interaction Between Domains of Strength and Weakness. *Cortex, 40*(1), 85-101.

Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory, 7*(2), 209-231.

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *Q J Exp Psychol A, 54*(2), 579-597. doi: 10.1080/713755977 [doi]

Piaget, J. (1976). The Hanoi Tower *The grasp of consciousness: action and concept in the young child* (pp. 287-299). Cambridge, Mass.: Harvard University Press.

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children Manual*. London: The Psychological Corporation.

Pinter, J., Brown, W. E., Eliez, S., Schmitt, J., Capone, G., & Reiss, A. (2001). Amygdala and hippocampal volumes in children with Down syndrome: a high-resolution MRI study. *Neurology, 56*(7), 972-974.

Pitcairn, T. K., & Wishart, J. G. (1994). Reactions of young children with Down's syndrome to an impossible task. *British Journal of Developmental Psychology, 12*(4), 485-489. doi: 10.1111/j.2044-835X.1994.tb00649.x

Pober, B. R. (2010). Williams–Beuren syndrome. *New England Journal of Medicine, 362*(3), 239-252.

Porter, M. A., Coltheart, M., & Langdon, R. (2007). The neuropsychological basis of hypersociability in Williams and Down syndrome. *Neuropsychologia, 45*(12), 2839-2849.

Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 3-30). Cambridge: Cambridge University Press.

Problem-Solving Style Questionnaire (PSSQ). (n.d.). Retrieved 25.06.13, 2013, from www.uk.sagepub.com/.../Problem-solving%20Style%20Questionnaire.doc

Puustinen, M., Lyyra, A.-L., Metsäpelto, R.-L., & Pulkkinen, L. (2008). Children's help seeking: The role of parenting. *Learning and Instruction, 18*(2), 160-171. doi: http://dx.doi.org/10.1016/j.learninstruc.2007.02.001

Rabbitt, P. (1997). Methodology of frontal and executive function

Raven, J. (2004). *Raven's Coloured Progressive Matrices*. London: Pearson.

Raven, J. C., Court, J. H., & Raven, J. (1990). *Coloured progressive matrices*. Oxford: Oxford Psychologists Press.

Reiss, A. L., Eckert, M. A., Rose, F. E., Karchemskiy, A., Kesler, S., Chang, M., . . . Galaburda, A. (2004). An experiment of nature: brain anatomy parallels cognition and behavior in Williams syndrome. *The Journal of Neuroscience, 24*(21), 5009-5015.

Reiss, A. L., Eliez, S., Schmitt, J. E., Straus, E., Lai, Z., Jones, W., & Bellugi, U. (2000). IV. Neuroanatomy of Williams syndrome: a high-resolution MRI study. *Journal of Cognitive Neuroscience, 12*(Supplement 1), 65-73.

Rhodes, S. M., Riby, D. M., Fraser, E., & Campbell, L. E. (2011). The extent of working memory deficits associated with Williams syndrome: Exploration of verbal and spatial domains and executively controlled processes. *Brain and Cognition, 77*(2), 208-214.

Rhodes, S. M., Riby, D. M., Park, J., Fraser, E., & Campbell, L. E. (2010). Executive neuropsychological functioning in individuals with Williams syndrome. *Neuropsychologia, 48*(5), 1216-1226.

Riby, D. M., & Hancock, P. J. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia, 46*(11), 2855-2860.

Riccio, C. A., Wolfe, M. E., Romine, C., Davis, B., & Sullivan, J. R. (2004). The Tower of London and neuropsychological assessment of ADHD in adults. *Archives of Clinical Neuropsychology, 19*(5), 661-671. doi: http://dx.doi.org/10.1016/j.acn.2003.09.001

Robertson, S. I. (2001). *Problem solving*. Hove: Psychology.

Rogers, S., Hepburn, S., & Wehner, E. (2003). Parent Reports of Sensory Symptoms in Toddlers with Autism and Those with Other Developmental Disorders. *Journal of Autism and Developmental Disorders, 33*(6), 631-642. doi: 10.1023/B:JADD.0000006000.38991.a7

Roid, G. H. (2003). *Stanford–Binet intelligence scales* (5th ed.). Itasca, IL: RiversidePublishing.

Roid, G. H., & Miller, L. J. (2002). *Leiter International Performance Scale-Revised (Leiter-R)*: Wood Dale, IL.

Roid, R., & Miller, J. (2002). *Leiter – R. International Performance Scale – Revised*: Giunti OS.

Ronnlund, M., Lovden, M., & Nilsson, L. G. (2001). Adult age differences in tower of Hanoi performance: Influence from demographic and cognitive variables. *Aging Neuropsychology and Cognition, 8*(4), 269-283.

Rosner, B., Hodapp, R., Fidler, D., Sagun, J., & Dykens, E. (2004). Social Competence in Persons with Prader-Willi, Williams and Down's Syndromes. *Journal of Applied Research in Intellectual Disabilities, 17*(3), 209-217. doi: 10.1111/j.1468-3148.2004.00200.x

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103*(1), 55.

Rowe, J., Lavender, A., & Turk, V. (2006). Cognitive executive function in Down's syndrome. *British Journal of Clinical Psychology, 45*(1), 5-17. doi: 10.1348/014466505x29594

Rowe, J. B., Owen, A. M., Johnsrude, I. S., & Passingham, R. E. (2001). Imaging the mental components of a planning task. *Neuropsychologia, 39*(3), 315-327.

Sahakian, B. J., & Owen, A. M. (1992). Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *Journal of the Royal Society of Medicine, 85*(7), 399.

Sampaio, A., Sousa, N., Férnandez, M., Henriques, M., & Gonçalves, O. F. (2008). Memory abilities in Williams syndrome: Dissociation or developmental delay hypothesis? *Brain and Cognition, 66*(3), 290-297.

Schmitt, J. E., Eliez, S., Warsofsky, I. S., Bellugi, U., & Reiss, A. L. (2001). Corpus callosum morphology of Williams syndrome: relation to genetics and behavior. *Developmental Medicine & Child Neurology, 43*(3), 155-159.

Schmitt, J. E., Watts, K., Eliez, S., Bellugi, U., Galaburda, A. M., & Reiss, A. L. (2002). Increased gyrification in Williams syndrome: evidence using 3D MRI methods. *Developmental Medicine & Child Neurology, 44*(5), 292-295.

Schnirman, G. M., Welsh, M. C., & Retzlaff, P. D. (1998). Development of the Tower of London-revised. *Assessment, 5*(4), 355-360.

Schunn, C. D., & Klahr, D. (1998). Production systems. In W. Bechtel & G. Graham (Eds.), A companion to cognitive science. Oxford, UK: Blackwell Publishing Ltd.

Semel, E., & Rosner, S. R. (2003). *Understanding Williams Syndrome: Behavioral Patterns and Interventions*. Taylor & Francis e-Library, 2011.: Lawrence Erlbaum Associates Inc.

Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., & Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion, 6*(1), 40.

Shallice, T. (1982). Specific Impairments of Planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 298*(1089), 199-209.

Shallice, T., & Burgess, P. (1991a). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*(2), 727-741.

Shallice, T., & Burgess, P. (1991b). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125–138). Oxford: Oxford University Press.

Siegler, R. S. (1983). Five generalizations about cognitive development. *American Psychologist, 38*(3), 263.

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking*. Indiana, USA: Prentice Hall.

Silverman, W. K., & Albano, A. M. (1996). *Anxiety Disorders Interview Schedule for DSM-IV.: Parent interview schedule* (Vol. 1): Oxford University Press.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*(2), 268-288.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145-159.

Singer Harris, N. G., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. *Developmental Neuropsychology, 13*(3), 345-370.

Spitz, H. H., Minsky, S. K., & Bessellieu, C. L. (1985). Influence of planning time and 1st-move strategy on tower of hanoi problem-solving performance of mentally-retarded young-adults and nonretarded-children. *American Journal of Mental Deficiency, 90*(1), 46-56.

Spitz, H. H., Webster, N. A., & Borys, S. V. (1982). Further studies of the Tower of Hanoi problem-solving performance of retarded young adults and nonretarded children. *Developmental Psychology, 18*(6), 922-930.

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology, 21*(4), 327-337. doi: http://dx.doi.org/10.1016/j.acn.2006.04.004

Stinton, C., Elison, S., & Howlin, P. (2010). Mental health problems in adults with Williams syndrome. *American Journal on Intellectual and Developmental Disabilities, 115*(1), 3-18.

Stinton, C., & Howlin, P. (2012). Adult outcomes and integration into society. In E. K. Farran & A. Karmiloff-Smith (Eds.), *Neurodevelopmental Disorders Across the Lifespan: A Neuroconstructivist Approach*. Oxford: Oxford University Press.

Stojanovik, V. (2006). Social interaction deficits and conversational inadequacy in Williams syndrome. *Journal of Neurolinguistics, 19*(2), 157-173.

Strømme, P., Bjømstad, P. G., & Ramstad, K. (2002). Prevalence estimation of Williams syndrome. *Journal of child neurology, 17*(4), 269-271.

Stroop, J. R. (1935). Studies of inference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Stuss, D. T. (1987). Contribution of frontal lobe injury to cognitive impairment after closed head injury: methods of assessment and recent findings. *Neurobehavioral recovery from head injury*, 166-177.

Stuss, D. T. (1992). Biological and psychological development of executive functions. *Brain and Cognition, 20*(1), 8-23.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological Research, 63*(3), 289-298. doi: 10.1007/s004269900007

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95*(1), 3-28. doi: 10.1037/0033-2909.95.1.3

Suchman, L. (1993). Response to Vera and Simon's Situated Action: A Symbolic Interpretation. *Cognitive Science, 17*(1), 71-75. doi: 10.1207/s15516709cog1701_5

Szepkouski, G. M., Gauvain, M., & Carberry, M. (1994). The development of planning skills in children with and without mental retardation. *Journal of Applied Developmental Psychology, 15*(2), 187-206.

Tager-Flusberg, H., Plesa-Skwerer, D., Faja, S., & Joseph, R. M. (2003). People with Williams syndrome process faces holistically. *Cognition, 89*(1), 11-24.

Tager-Flusberg, H., Sullivan, K., & Boshart, J. (1997). Executive functions and performance on false belief tasks. *Developmental Neuropsychology, 13*(4), 487-493. doi: 10.1080/87565649709540689

Temple, C. M., Carney, R. A., & Mullarkey, S. (1996). Frontal lobe function and executive skills in children with Turner's syndrome. *Developmental Neuropsychology, 12*(3), 343-363.

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and*

Hearing Research, 52(2), 336-358. doi: 1092-4388_2009_07-0144 [pii]10.1044/1092-4388(2009/07-0144) [doi]

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2008). Executive Functions: Performance-Based Measures and the Behavior Rating Inventory of Executive Function (BRIEF) in Adolescents with Attention Deficit/Hyperactivity Disorder (ADHD). *Child Neuropsychology, 15*(1), 53-72. doi: 10.1080/09297040802070929

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54*(2), 131-143.

Tsao, R., & Kindelberger, C. (2009). Variability of cognitive development in children with Down syndrome: Relevance of good reasons for using the cluster procedure. *Research in Developmental Disabilities, 30*(3), 426-432.

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology-Paris, 99*(4-6), 308-317.

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., . . . Halsband, U. (2004). Planning Abilities and the Tower of London: Is This Task Measuring a Discrete Cognitive Function? *Journal of Clinical and Experimental Neuropsychology, 26*(6), 846-856. doi: 10.1080/13803390490509574

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: the impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research, 17*(3), 675-683. doi: http://dx.doi.org/10.1016/S0926-6410(03)00191-5

Vakil, E., & Lifshitz-Zehavi, H. (2012). Solving the Raven Progressive Matrices by adults with intellectual disability with/without Down syndrome: Different cognitive patterns as indicated by eye-movements. *Research in Developmental Disabilities, 33*(2), 645-654. doi: http://dx.doi.org/10.1016/j.ridd.2011.11.009

Van Herwegen, J., Farran, E., & Annaz, D. (2011). Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome. *Research in Developmental Disabilities, 32*(1), 93-99. doi: http://dx.doi.org/10.1016/j.ridd.2010.09.005

van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science, 27*(2), 227-257. doi: http://dx.doi.org/10.1016/S0364-0213(02)00116-7

Vaus, D. d. (2002). Constructing Questionnaires *Surveys in Social Research* (5th ed.). NSW, Australia Allen & Unwin.

Venuti, P., de Falco, S., Esposito, G., Zaninelli, M., & Bornstein, M. H. (2012). Maternal functional speech to children: A comparison of autism spectrum disorder, Down syndrome, and typical development. *Research in Developmental Disabilities, 33*(2), 506-517. doi: http://dx.doi.org/10.1016/j.ridd.2011.10.018

Vera, A. H., & Simon, H. A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science, 17*(1), 7-48. doi: 10.1207/s15516709cog1701_2

Vianello, R., & Marin, M. L. (1997). *OLC: Operazioni Logiche e Conservazione. Progetto MS.* Bergamo: Junior.

Vicari, S. (2006). Motor Development and Neuropsychological Patterns in Persons with Down Syndrome. *Behavior Genetics, 36*(3), 355-364. doi: 10.1007/s10519-006-9057-8

Vicari, S., Bellucci, S., & Carlesimo, G. A. (2000). Implicit and explicit memory: a functional dissociation in persons with Down syndrome. *Neuropsychologia, 38*(3), 240-251.

Vicari, S., Bellucci, S., & Carlesimo, G. A. (2001). Procedural learning deficit in children with Williams syndrome. *Neuropsychologia, 39*(7), 665-677.

Vicari, S., Bellucci, S., & Carlesimo, G. A. (2005). Visual and spatial long-term memory: differential pattern of impairments in Williams and Down syndromes. *Developmental Medicine & Child Neurology 47*, 305-311: Cambridge Journals Online.

Vicari, S., Carlesimo, A., & Caltagirone, C. (1995). Short-term memory in persons with intellectual disabilities and Down's syndrome. *Journal of Intellectual Disability Research, 39*(6), 532-537.

Vicari, S., & Carlesimo, G. A. (2006). Short-term memory deficits are not uniform in Down and Williams syndromes. *Neuropsychology review, 16*(2), 87-94.

von Arnim, G., & Engel, P. (1964). Mental Retardation Related to Hypercalcaemia. *Developmental Medicine & Child Neurology, 6*(4), 366-377. doi: 10.1111/j.1469-8749.1964.tb08138.x

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology, 16*(2), 317 - 322.

Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc Tower of London task. *The Quarterly Journal of Experimental Psychology Section A, 50*(1), 49-78.

Wechsler, D., & Naglieri, J. (2006). *Wechsler Nonverbal scale of ability*. USA: Harcourt Assessment.

Weis, S. (1991). Morphometry and magnetic resonance imaging of the human brain in normal controls and Down's syndrome. *The Anatomical Record, 231*(4), 593-598. doi: 10.1002/ar.1092310421

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the tower of hanoi disk-transfer task. *Cognitive Development, 6*(1), 59-76.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of Working Memory and Inhibition to Performance. *Brain and Cognition, 41*(2), 231-242.

Willner, P., Bailey, R., Parry, R., & Dymond, S. (2010). Evaluation of executive functioning in people with intellectual disabilities. *Journal of Intellectual Disability Research, 54*(4), 366-379.

Wilson, B., Cockburn, J., & Baddeley, A. (1985). The Rivermead behavioural memory test manual. *Bury St Edmunds: Thames Valley Test Co*.

Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation, 8*(3), 213-228.

Wing, L. (1980). The MRC Handicaps, Behaviour & Skills (HBS) schedule. *Acta Psychiatrica Scandinavica, 62*(S285), 241-248. doi: 10.1111/j.1600-0447.1980.tb07696.x

Winsler, A., & Naglieri, J. (2003). Overt and Covert Verbal Problem-Solving Strategies: Developmental Trends in Use, Awareness, and Relations With Task Performance in Children Aged 5 to 17. *Child Development, 74*(3), 659-678. doi: 10.1111/1467-8624.00561

Wishart, J. (1993a). The development of learning difficulties in children with Down's syndrome. *Journal of Intellectual Disability Research, 37*(4), 389-403.

Wishart, J. (1993b). Learning the hard way: Avoidance strategies in young children with Down syndrome. *Down Syndrome Research and Practice, 1*(2), 47-55.

Woodruff-Borden, J., Kistler, D. J., Henderson, D. R., Crawford, N. A., & Mervis, C. B. (2010). Longitudinal course of anxiety in children and adolescents with Williams syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 154C*(2), 277-290. doi: 10.1002/ajmg.c.30259

Wuang, Y. P., & Su, C. Y. (2011). Correlations of sensory processing and visual organization ability with participation in school-aged children with Down syndrome. *Res Dev Disabil, 32*(6), 2398-2407. doi: S0891-4222(11)00277-0 [pii]10.1016/j.ridd.2011.07.020 [doi]

Young, R. M. (2001). Production Systems in Cognitive PsychologyInternational Encyclopedia of the Social & Behavioral Sciences.

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297-301.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early Development of Executive Function: A Problem-Solving Framework. *Review of General Psychology, 1*(2), 198-226.

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., . . . Carlson, S. M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development, 68*(3), i-151.

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition, 56*(3), 286-292.

APPENDICES

Appendix A

Tower of London Problem Set

Minimum Moves	Start State			Goal State			Berg &Byrd (2002) notation	Goal Configuration PT = partial tower	Search Depth	Intermediate moves	Counterintuitive moves (further from or off goal peg)	Number of optimal paths	Suboptimal Paths: Detour/ Dead End
	L	C	R	L	C	R							
1 (D)		wr	b	r	wb		13-14	PT	0	0	0	1	0
1 (P1)		br	w	w	br		46-54	PT	0	0	0	1	0
2 (P2)	br		w	r	bw		22-24	PT	0	0	0	1	0
1	r	wb			wb	r	14-66	PT	0	0	0	1	0
1	b	rw			rw	b	34-26	PT	0	0	0	1	0
1	w	br			br	w	54-46	PT	0	0	0	1	0
1	b	wr			wr	b	44-56	PT	0	0	0	1	0
2		wb	r	rw	b		66-63	PT	0	0	0	1	0
2		rw	b	rb	w		26-33	PT	0	0	0	1	0
2		br	w	wb	r		46-43	PT	0	0	0	1	0
2		bw	r	br	w		36-23	PT	0	0	0	1	0
3	br	w			rb	w	23-16	PT	0	0	0	1	1 detour
3	wb	r			bw	r	43-36	PT	0	0	0	1	1 detour
4	rw	b		b	wr		63-44	PT	1	1	1	1	1 detour
4	br	w		w	rb		23-64	PT	1	1	1	1	1 detour
5	wr	b		bw	r		13-53	PT	2	2	1	1	1 detour
5	rb	w		wr	b		33-13	PT	2	2	1	1	1 detour
6		wb	r	b	rw		66-34	PT	3	3	2	1	5 detours
6		wr	b	r	bw		56-24	PT	3	3	2	1	5 detours

Table A-1: Tower of London problem set. Letters denote the position of the pieces such that , for example, rw = red placed on top of white. PT = partial tower (goal state). L, C and R refer to the peg position: L = left, C = centre, R = right. D = demonstration trial, P1 = first practice trial, P2 = second practice trial.

Appendix B

Chapter 2: Developmental trajectories with significant relationships in one group only

Variables that were only significantly related to the TOL score in one group were analysed using the developmental trajectory approach, and are summarised in Table B-1 below. No trajectory analyses were conducted for chronological age (CA) due to the lack of overlap between the TD group and the atypical group scores, and the lack of relationship between CA and TOL score in the atypical groups. Trajectory analysis was not conducted for the forwards block and digit span measures due to non-normal data which were not improved by transformations. Details of correlations between measures and TOL score can be found in Table 2.5. See figures B-1 to B-6 below for scatter plots of these data.

Variable	N (TD, DS, WS)	Group effect	Covariate effect	Group by Covariate interaction	*Post hoc* ANCOVA
RCPM	56, 20, 20	F(2,90) = 1.986, p = .143, partial η^2 = .042	F(1,90) = 8.887, p = .004, partial η^2 = .090	p = .643	N/A
Inhibition % Opposite[12]	56, 20, 20	F(2,90) = 2.418, p = .095, partial η^2 = .051	F(1,90) = 9.411, p = .003, partial η^2 = .095	F<1	TD < WS DS = TD,WS
Shifts	54, 18, 20	F(2,86) = 3.514, p = .034, partial η^2 = .076	F(1,86) = 9.004, p = .004, partial η^2 = .095	F<1	TD < WS DS = WS,TD

Table continued overleaf

[12] Upon excluding data from one participant in the WS group with a low score on accuracy in the Opposite condition, the correlation in this group loses significance (r = -.340, p = .155), leaving none of the three groups with a significant correlation between accuracy in the opposite condition and TOL score. ANCOVA without this participant's data no longer indicates a Group effect (F(2,89) = 2.024, p = .138, partial η^2 = .044).

Table B-1 continued

Variable	N (TD, DS, WS)	Group effect	Covariate effect	Group by Covariate interaction	*Post hoc* ANCOVA
Inhibition: RT Change	56, 20, 20	$F(2,90) = 10.373$, $p < .001$, partial $\eta^2 = .187$	$F(1,90) = 3.517$, $p = .064$, partial $\eta^2 = .038$	$p = .286$	TD < DS,WS

Table B- 1: Outcomes of developmental trajectory ANCOVAs for significant TOL associations for one group only. In the *post hoc* ANCOVA column, note that '<' indicates lower scores on the TOL, i.e., better performance.

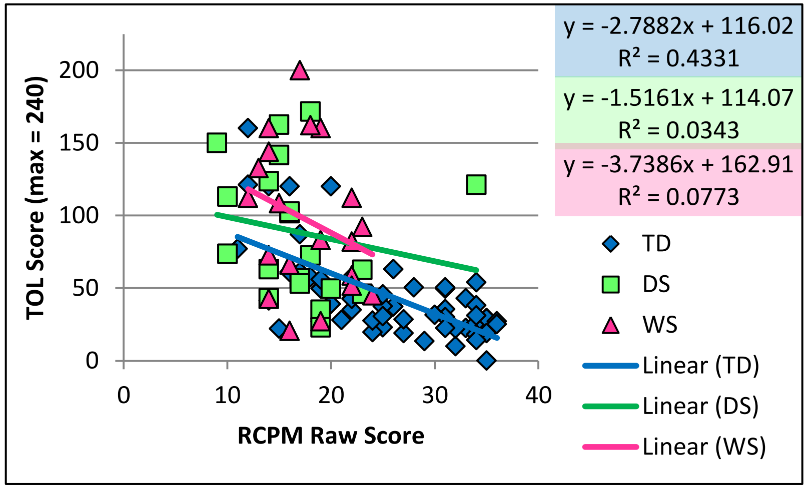

Figure B- 1: Scatter plot of RCPM and TOL score by group

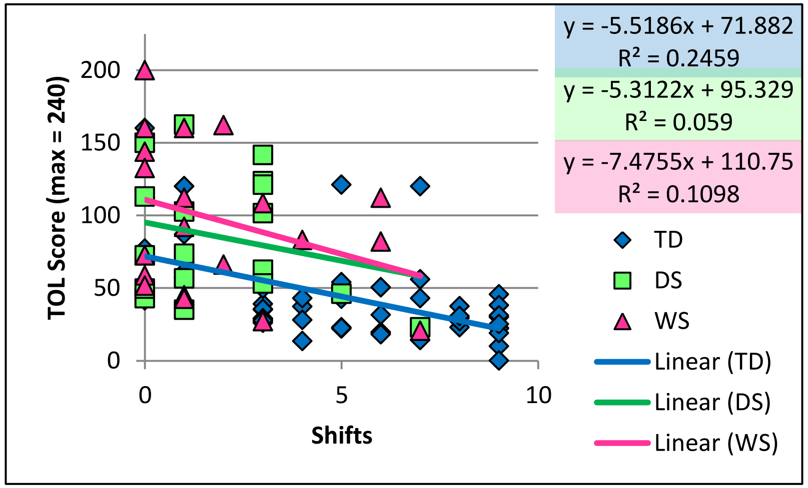

Figure B- 2: Scatter plot of shifts made and TOL score by group

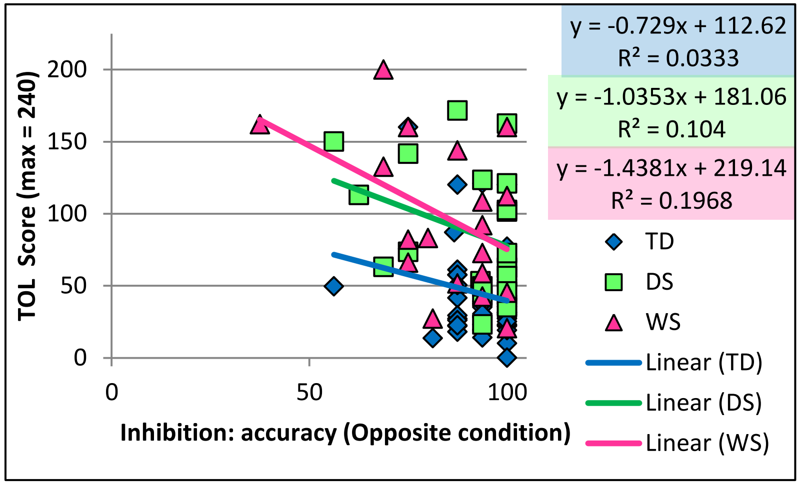

Figure B- 3: Scatter plot of inhibition accuracy (Opposite) and TOL score by group

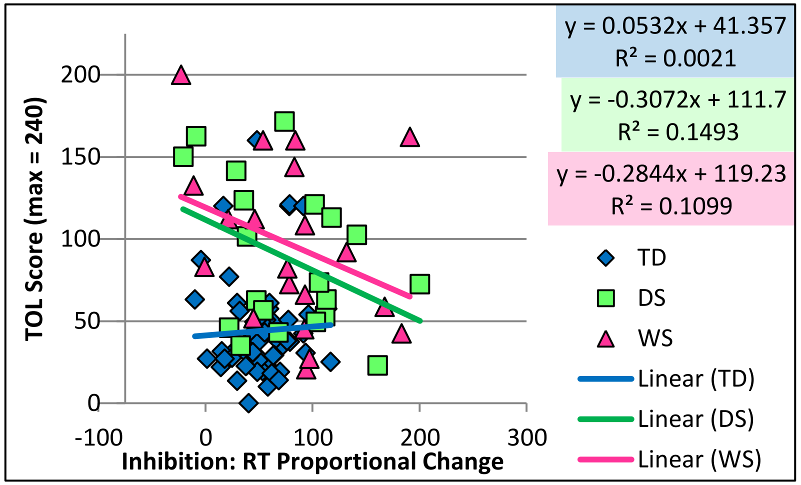

Figure B -4: Scatter plot of inhibition RT proportional change and TOL score by group

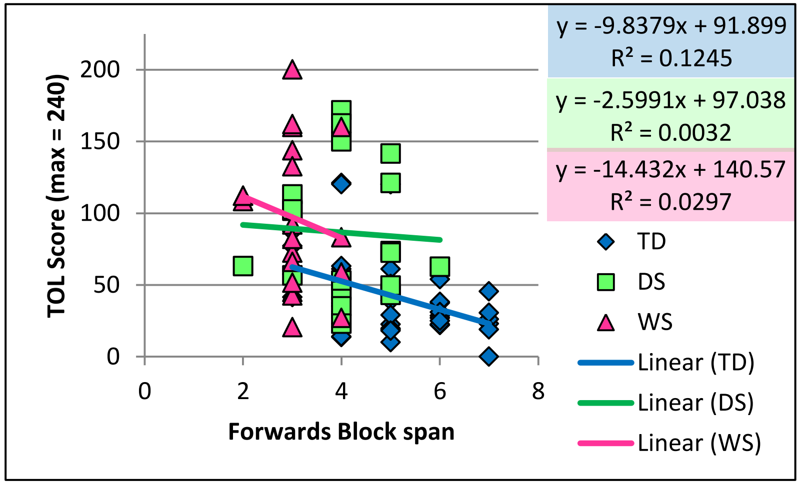

Figure B- 5: Scatter plot of forwards block span and TOL score by group

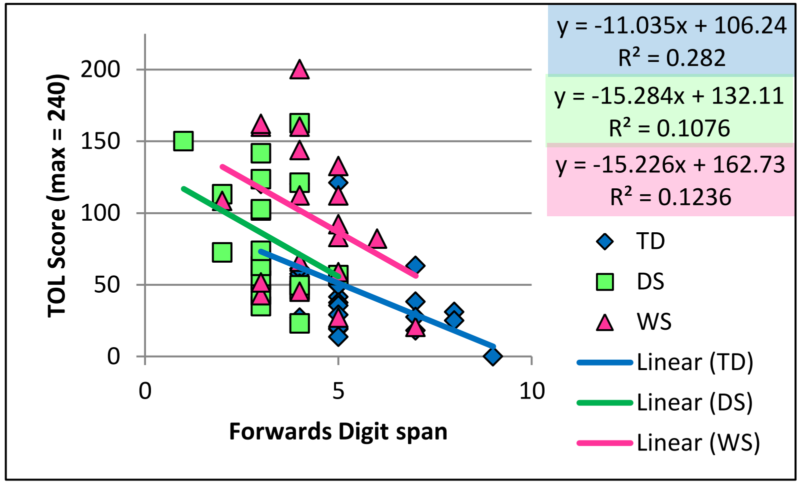

Figure B- 6: Scatter plot of forwards digit span and TOL score by group

Appendix C

BRIEF Example Items

Item 28: Gets caught up in details and misses the big picture (Plan/Organize)

Item 55: Has trouble putting the brakes on his/her actions (Inhibit)

Item 68: Leaves a trail of belongings wherever he/she goes (Organization of Materials)

Appendix D

The Problem-Solving Questionnaire

In this section, the PSQ questions are displayed (with adjusted margins and spacing for inclusion in this thesis).

Part A: Everyday Tasks (1): Getting dressed

1a. Please indicate whether you son or daughter usually gets dressed **independently (1)** with **some help (2)** or only if **someone else dresses them (3)** by circling the relevant number below:

(please circle one only) **1 2 3**

1b. If you circled with **some help (2)** above, please indicate whether help is in the form of: ***(please tick as many as apply)***

Specific instructions (e.g., "now you need to…"), ☐

General encouragement (e.g., "well done, keep going, nearly there…") ☐

Physical support (e.g. someone holding objects still or doing part(s) of the task for them) ☐

1c. Who usually selects the clothes to be worn by your son or daughter? ***(please tick one only)***

They select the clothes **themselves** ☐ **Someone else** selects clothes for them ☐

1d. For each statement below, please indicate how easy your son or daughter finds each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				**Very Difficult**
i. Knowing what the goal of the problem is (being fully dressed)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what can be done to reach the goal (putting clothes on)	**1**	**2**	**3**	**4**	**5**
iii. Doing steps in the right order (e.g. putting on socks before shoes)	**1**	**2**	**3**	**4**	**5**
iv. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
v. Stopping when a task is finished (e.g. stopping looking for clothes in the wardrobe when fully dressed)	**1**	**2**	**3**	**4**	**5**
vi. Finishing off the task (e.g. closing the wardrobe door)	**1**	**2**	**3**	**4**	**5**

1e. **If something goes wrong** (for example a t shirt is put on backwards), please indicate how likely your son or daughter

is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

i. Recognise that there is a problem (e.g. t shirt is on backwards)	**1**	**2**	**3**	**4**	**5**
ii. Change what they have been doing in response to the problem (e.g. try to turn the t shirt round)	**1**	**2**	**3**	**4**	**5**
iii. Ask for help	**1**	**2**	**3**	**4**	**5**
iv. Become emotional	**1**	**2**	**3**	**4**	**5**
v. Lose focus	**1**	**2**	**3**	**4**	**5**
vi. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vii. Reach the solution	**1**	**2**	**3**	**4**	**5**

viii. Are they likely to do something else? ***(please explain here)***

Everyday Tasks (2): Brushing Teeth

2a. Please indicate whether your son or daughter usually brushes their teeth **independently (1)** with **some help (2)** or only if **someone else brushes them (3)** by circling the relevant number below:

(please circle one only) **1** **2** **3**

2b. If you circled with **some help (2)** above, please indicate whether help is in the form of: ***(please tick as many as apply)***

Specific instructions (e.g., "now you need to…"), ☐

General encouragement (e.g., "well done, keep going, nearly there…") ☐

Physical support (e.g. someone holding objects still or doing part(s) of the task for them) ☐

2c. For each statement below, please indicate how easy your son or daughter finds each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.	**Very Easy**				**Very Difficult**
i. Knowing what the goal of the problem is (having clean teeth)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what can be done to reach the goal (brushing teeth)	**1**	**2**	**3**	**4**	**5**
iii. Doing steps in the right order (e.g. putting toothpaste on the brush before starting to brush)	**1**	**2**	**3**	**4**	**5**
iv. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**

v. Stopping when a task is finished (e.g. teeth are clean)	**1**	**2**	**3**	**4**	**5**
vi. Finishing off the task (e.g. putting the toothbrush and toothpaste away after brushing)	**1**	**2**	**3**	**4**	**5**

2d. **If something goes wrong** (for example toothpaste is forgotten), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				**Very Unlikely**
i. Recognise that there is a problem (e.g. toothpaste has been forgotten)	**1**	**2**	**3**	**4**	**5**
ii. Change what they have been doing in response to the problem (e.g. attempt to add toothpaste)	**1**	**2**	**3**	**4**	**5**
iii. Ask for help	**1**	**2**	**3**	**4**	**5**
iv. Become emotional	**1**	**2**	**3**	**4**	**5**
v. Lose focus	**1**	**2**	**3**	**4**	**5**
vi. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vii. Reach the solution	**1**	**2**	**3**	**4**	**5**

viii. Are they likely to do something else? ***(please explain here)***

Everyday Tasks (3): Making a sandwich

3a. Please indicate how often your son or daughter on average attempts to make a sandwich: **every day or more often (1), a few times a week (2), a few times a month (3), a few times a year (4), or never (5),** by circling the relevant number below:
(please circle one only) **1 2 3 4 5**

3b. Now please indicate whether you son or daughter usually makes a sandwich **independently (1)** with **some help (2)** or only has a sandwich if **someone else makes it for them (3)** by circling the relevant number below:
(please circle one only) **1 2 3**

3c. If you circled with **some help (2)** above, please indicate whether help is in the form of: ***(please tick as many as apply)***

Specific instructions (e.g., “now you need to…”), ☐

☐

General encouragement (e.g., “well done, keep going, nearly there…”)

Physical support (e.g. someone holding objects still or doing part(s) of the task for them) ☐

3d. For each statement below, please indicate how easy your son or daughter finds each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				Very Difficult
i. Knowing what the goal of the problem is (making the sandwich)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what can be done to reach the goal (putting the filling on the bread; cutting the sandwich)	**1**	**2**	**3**	**4**	**5**
iii. Doing steps in the right order (e.g. starting with bread, then filling, then more bread)	**1**	**2**	**3**	**4**	**5**
iv. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
v. Stopping when a task is finished (e.g. when the sandwich is made)	**1**	**2**	**3**	**4**	**5**
vi. Finishing off the task (e.g. putting the ingredients and plate away afterwards)	**1**	**2**	**3**	**4**	**5**

3e. **If something goes wrong** (for example the filling is forgotten), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				Very Unlikely
i. Recognise that there is a problem (e.g. no filling is in the sandwich)	**1**	**2**	**3**	**4**	**5**
ii. Change what they have been doing in response to the problem (e.g. take the top piece of bread away and add a filling)	**1**	**2**	**3**	**4**	**5**
iii. Ask for help	**1**	**2**	**3**	**4**	**5**
iv. Become emotional	**1**	**2**	**3**	**4**	**5**
v. Lose focus	**1**	**2**	**3**	**4**	**5**
vi. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vii. Reach the solution	**1**	**2**	**3**	**4**	**5**

viii. Are they likely to do something else? ***(please explain here)***

Everyday Tasks (4): Making a telephone call

4a. Please indicate how often your son or daughter on average attempts to make a telephone call: **every day or more often (1), a few times a week (2), a few times a month (3), a few times a year (4), or never (5),** by circling the relevant number below:

(please circle one only) **1 2 3 4 5**

4b. Now please indicate whether you son or daughter usually makes phone calls **independently (1)** with **some help (2)** or only if **someone else calls for them (3)** by circling the relevant number below:

(please circle one only) **1 2 3**

4c. If you circled with **some help (2)** above, please indicate whether help is in the form of: ***(please tick as many as apply)***

Specific instructions (e.g., "now you need to…"), ☐

General encouragement (e.g., "well done, keep going, nearly there…") ☐

Physical support (e.g. someone holding objects still or doing part(s) of the task for them) ☐

4d. For each statement below, please indicate how easy your son or daughter finds each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				Very Difficult
i. Knowing what the goal of the problem is (for example making a call to ask someone a question)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what can be done to reach the goal (dialling, speaking)	**1**	**2**	**3**	**4**	**5**
iii. Doing steps in the right order (e.g. waiting for the dial tone before dialling)	**1**	**2**	**3**	**4**	**5**
iv. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
v. Stopping when a task is finished (e.g. hanging up when call is complete)	**1**	**2**	**3**	**4**	**5**

4e. **If something goes wrong** (for example the wrong number is dialled), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				Very Unlikely
i. Recognise that there is a problem (e.g. call placed to the wrong number)	**1**	**2**	**3**	**4**	**5**
ii. Change what they have been doing in response to the problem (e.g. try a different number)	**1**	**2**	**3**	**4**	**5**

iii. Ask for help	**1**	**2**	**3**	**4**	**5**
iv. Become emotional	**1**	**2**	**3**	**4**	**5**
v. Lose focus	**1**	**2**	**3**	**4**	**5**
vi. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vii. Reach the solution	**1**	**2**	**3**	**4**	**5**

viii. Are they likely to do something else? *(please explain here)*

Everyday Tasks: Strategies

Finally, if there are any strategies that you or your son or daughter uses to help with everyday tasks, we would really like to know about them. As an example, a favourite rhyme or set of pictures might help them remember what needs to be done in a task. The following table can be used to write down the strategies. Please also let us know whether these are strategies that were initiated by your son or daughter or whether they have been taught to use them. Don't feel that you are expected to fill in the whole table: the space is only there if you would like to use it! Thank you.

Task	Strategy	Was the strategy taught or self-initiated?

Part B: Novel Problems (1): Finding a lost possession

Sometimes we encounter a problem in everyday life that is unexpected or is not something we routinely do every day. The next few questions give some examples of some potentially novel problems that might still be within the realm of everyday life. The first is **finding a lost possession**. Please imagine that your son or daughter has mislaid an everyday item such as a set of keys, wallet or favourite book or toy (depending on their age).

5a. For each statement below, please indicate how easy your son or daughter would find each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				**Very Difficult**
i. Recognising that there is a problem (item is lost)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what the goal of the problem is (finding the item)	**1**	**2**	**3**	**4**	**5**
iii. Having some ideas about what could be done to reach the goal (e.g. searching or asking someone)	**1**	**2**	**3**	**4**	**5**
iv. Keeping track of what they are doing (e.g. avoiding revisiting places they have already searched in)	**1**	**2**	**3**	**4**	**5**
v. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
vi. Use their existing knowledge to help them (e.g. looking for a book on a bookshelf or a wallet in a coat pocket)	**1**	**2**	**3**	**4**	**5**
vii. Applying a strategy they have previously learnt in another situation	**1**	**2**	**3**	**4**	**5**
viii. Stopping when a task is finished (item is found)	**1**	**2**	**3**	**4**	**5**

5b. **If something goes wrong** (for example searching in one location is fruitless), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				**Very Unlikely**
i. Change what they have been doing in response to the problem (e.g. searching somewhere else or in a different way)	**1**	**2**	**3**	**4**	**5**
ii. Ask for help	**1**	**2**	**3**	**4**	**5**
iii. Become emotional	**1**	**2**	**3**	**4**	**5**
iv. Lose focus	**1**	**2**	**3**	**4**	**5**
v. End their search after a sensible amount of time if the item has not been found	**1**	**2**	**3**	**4**	**5**

vi. End their search through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vii. Reach the solution (find the item)	**1**	**2**	**3**	**4**	**5**

viii. Are they likely to do something else? ***(please explain here)***

5c**. If your son or daughter fails to find the item, what do you think the reason for this is most likely to be?**

Novel Problems (2): Packing a bag for the day

What might need to be packed in a bag each day varies depending on what is planned for that day, the weather, and so on. This question is about how your son or daughter tends to respond to these changing demands.

6a. For each statement below, please indicate how easy your son or daughter would find each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				**Very Difficult**
i. Recognising that there is a problem (bag is not yet packed)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what the goal of the problem is (having a bag packed with everything they will need)	**1**	**2**	**3**	**4**	**5**
iii. Having some ideas about what could be done to reach the goal (e.g. collecting items to pack)	**1**	**2**	**3**	**4**	**5**
iv. Keeping track of what they are doing (e.g. not searching for items they have already packed)	**1**	**2**	**3**	**4**	**5**
v. Being flexible in what they choose to pack, depending on what is needed, e.g. only choosing what they will need that day	**1**	**2**	**3**	**4**	**5**
vi. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
vii. Use their existing knowledge to help them (e.g. packing an umbrella if it is raining)	**1**	**2**	**3**	**4**	**5**
viii. Applying a strategy they have previously learnt in another situation	**1**	**2**	**3**	**4**	**5**
ix. Stopping when the task is finished (bag is packed)	**1**	**2**	**3**	**4**	**5**

6b. **If something goes wrong** (for example an item has been squashed in the bag), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				Very Unlikely
i. Change what they have been doing in response to the problem (e.g. repacking the bag)	**1**	**2**	**3**	**4**	**5**
ii. Ask for help	**1**	**2**	**3**	**4**	**5**
iii. Become emotional	**1**	**2**	**3**	**4**	**5**
iv. Lose focus	**1**	**2**	**3**	**4**	**5**
v. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vi. Reach the solution	**1**	**2**	**3**	**4**	**5**

vii. Are they likely to do something else? *(please explain here)*

6c**. If your son or daughter fails to pack the bag with what they will need, what do you think the reason for this is most likely to be?**

Novel Problems (3): Putting items away in a wardrobe/chest of drawers

Please consider times when your son or daughter may need to put items into a wardrobe or chest of drawers, for example when tidying their room.

7a. For each statement below, please indicate how easy your son or daughter would find each aspect of the task, **from 1 (very easy) to 5 (very difficult)** by circling the relevant number in each case.

	Very Easy				Very Difficult
i. Recognising that there is a problem (room is messy; hard to find things)	**1**	**2**	**3**	**4**	**5**
ii. Knowing what the goal of the problem is (that things are put away)	**1**	**2**	**3**	**4**	**5**
iii. Having some ideas about what could be done to reach the goal (e.g. putting things inside the chest of drawers)	**1**	**2**	**3**	**4**	**5**
iv. Keeping track of what they are doing (e.g. picking things up from one part of the room at a time)	**1**	**2**	**3**	**4**	**5**
v. Staying focused on the task (successfully ignoring distractions)	**1**	**2**	**3**	**4**	**5**
vi. Use their existing knowledge to help them (e.g. putting things in the	**1**	**2**	**3**	**4**	**5**

correct drawer, e.g. all the t shirts or games together)

vii. Applying a strategy they have previously learnt in another situation	**1**	**2**	**3**	**4**	**5**
viii. Stopping when a task is finished (items are put away successfully)	**1**	**2**	**3**	**4**	**5**

7b. **If something goes wrong** (for example a drawer will not close), please indicate how likely your son or daughter is to do the following things, **from 1 (very likely) to 5 (very unlikely)** by circling the relevant number in each case.

	Very Likely				**Very Unlikely**
i. Change what they have been doing in response to the problem (e.g. unpacking and repacking the drawer)	**1**	**2**	**3**	**4**	**5**
ii. Ask for help	**1**	**2**	**3**	**4**	**5**
iii. Become emotional	**1**	**2**	**3**	**4**	**5**
iv. Lose focus	**1**	**2**	**3**	**4**	**5**
v. Stop their efforts through lack of perseverance	**1**	**2**	**3**	**4**	**5**
vi. Reach the solution	**1**	**2**	**3**	**4**	**5**

vii. Are they likely to do something else? ***(please explain here)***

7c**. If your son or daughter fails to put the items away, what do you think the reason for this is most likely to be?**