

University of East London

Professional Doctorate in Clinical Psychology

Assessment Proforma

Please read the following candidate's declaration, and tick the adjacent boxes to confirm that you have complied with each statement. Then complete the cover sheet below in full. Failing to do will result in your assessment being returned to you for resubmission. Please raise any queries regarding this form with your academic tutor well in advance of submission.

CANDIDATE'S DECLARATION

I confirm that no part of this submitted material, except where clearly quoted and referenced, has been copied from material belonging to any other person (e.g., from a book, article, lecture handout, web site, or another candidate). I am aware that it is a breach of university regulations to copy the work of another without clear acknowledgement, and that attempting to do so will render me liable to disciplinary proceedings.

Tick to confirm

I confirm that the word count cited below is accurate, and within the limit allowed for this type of assessment. The count includes all words in the body of text, diagrams, tables and footnotes (though not the contents page, references or appendices). I have presented the assessed work with page margins, line spacing, font size and page numbers as required in the relevant section of the handbook and module guide.

Tick to confirm

COVER SHEET

UEL STUDENT NUMBER

U1438315

MODULE CODE

PY8205

WORK TO BE ASSESSED

(e.g., Year 1 Essay, Practical Report 3, SRR)

Thesis in partial fulfilment of
Professional Doctorate
(resubmission)

SUBMISSION DATE

7-08-2017

WORD COUNT

27102

**Investigating the use of board games as neuropsychological tests with
children with acquired brain injury**

Patrick Murphy

A thesis submitted in partial fulfilment of the requirements of the University of
East London for the degree of Professional Doctorate in Clinical Psychology

May 2017

Word Count: 27102

Abstract

This research project investigated the use of two board games, Guess Who and Connect 4, as initial neuropsychological assessment measures with children and young people. The validity of novel measures derived from game-play was investigated in a group of typically developing participants (N=14). The level of engagement offered by the games and the potential ecological validity of this assessment method was also investigated. This was to identify potential additional benefits to this method of assessment compared with traditional testing procedures. The performance of a small group of participants with acquired brain injury (N=5) was also explored to identify the potential of the novel measures to discern cognitive deficits in this group. As hypothesised, a measure of strategy derived from Guess Who demonstrated concurrent validity with two established measures of executive function, the D-KEFS Twenty Questions Test and the Zoo Map Test. Also in line with hypotheses, the number of wins identified on Connect 4 showed concurrent validity with a measure of visual search and attention, the Trail-Making Test Part A. The scores of the participants with acquired brain injury on the novel measures appeared weaker than the typically developing group, particularly for those scoring poorly on a measure of general non-verbal ability. Participants did not rate the games as any more engaging or any less anxiety provoking than the established measures. In-session observations during game-play did not shed light on the functional difficulties reported by parents on a standardised proxy report (the BRIEF). Alterations to game set-up that could increase the discriminant validity of the novel measures are discussed. The potential for this method to increase engagement in a less-high-functioning sample in acute care is also discussed. Issues with the “gamification” of assessment procedures are considered, including the difficulty in simultaneously gathering observational data and quantitative measures.

Acknowledgements

Thanks to my first supervisor, Dr. Jenny Jim, and second supervisor, Dr. Deirdre Birtles, for their guidance throughout the project and feedback on the drafts for this thesis.

Thanks to the staff at the site where the participants with acquired brain injury were recruited from. Their assistance with participant screening and recruitment and their comments on their own use of games during assessment and rehabilitation were invaluable.

Thanks to Dr. Howard Fine for his comments on his own use of the games in clinical work.

Thanks to my wife and brother-in-law who helped with the creation of the set of faces for Guess Who. Thanks also to friends and neighbours who played Guess Who with me early in the research process.

The biggest thank you goes to the participants, who took time away from study and leisure to participate and kindly let me into their homes to complete the study tasks.

Acronyms used in the thesis

| | |
|-------|--|
| ABI | Acquired Brain Injury |
| BRIEF | Behaviour Rating Inventory of Executive Function |
| CYP | Child or Young Person/Children or Young People |
| EE | Elimination Errors (on Guess Who) |
| GWSS | Guess Who Strategy Score |
| TD | Typically Developing |
| MCST | Modified Card Sorting Task |
| RPM | Raven's Standard Progressive Matrices |
| SS | Switching Score (on Connect 4) |
| TBI | Traumatic Brain Injury |
| TMT | Trail Making Test |
| WI | Wins Identified (on Connect 4) |
| WISC | Wechsler Intelligence Scale for Children |

| | |
|---|----|
| 1. Introduction | |
| 1.1 Study Overview..... | 8 |
| 1.2 ABI in CYP and the role of clinical psychology | 8 |
| 1.3 Issues with neuropsychological testing with CYP..... | 9 |
| 1.3.1 Developmental Considerations affecting test design..... | 10 |
| 1.3.2 ABI aetiology and screening test design..... | 11 |
| 1.3.3 Ecological Validity of executive function tests..... | 13 |
| 1.3.4 Engagement during testing..... | 15 |
| 1.3.5 Section summary..... | 16 |
| 1.4 Study aims: can board games be used as neuropsychological tests?..... | 17 |
| 1.4.1 Cognitive assessment via games or play..... | 17 |
| 1.4.2 Guess Who and Connect 4 as neuropsychological tests..... | 17 |
| 1.4.3 Potential psychometric and neuropsychological properties..... | 18 |
| 1.4.4 Observational data and ecological validity..... | 20 |
| 1.4.5 Guess Who and Connect 4 as screening tests..... | 21 |
| 1.5 Research Questions..... | 21 |
| 2. Methods | |
| 2.1 Epistemological framework and methodology..... | 22 |
| 2.2 Design..... | 23 |
| 2.3 Participants: Selection, recruitment and power analysis..... | 24 |
| 2.4 Guess Who and Connect Four: test development..... | 26 |
| 2.4.1 General description of games..... | 26 |
| 2.4.2 How could these games be used as tests?..... | 26 |
| 2.4.3 Test Development: Guess Who..... | 28 |
| 2.4.3.1 Creating a character set..... | 28 |
| 2.4.3.2 Game-play strategies..... | 28 |
| 2.4.3.3 Novel Guess Who measures: scoring procedure..... | 33 |
| 2.4.4 Test Development: Connect Four..... | 33 |
| 2.4.4.1 Establishing what constitutes a “good” performance..... | 33 |
| 2.4.4.2 Preliminary definitions..... | 34 |
| 2.4.4.3 A heuristic for Connect Four..... | 34 |
| 2.4.4.4 Novel Connect 4 measures: scoring procedure..... | 37 |

| | |
|--|----|
| 2.5 Neuropsychological battery details and test instructions..... | 39 |
| 2.5.1 Respondent burden..... | 39 |
| 2.5.2 Games and novel measures..... | 39 |
| 2.5.3 Battery of established measures..... | 41 |
| 2.6 Ethics..... | 50 |
| 2.7 General procedure..... | 51 |
| 2.8 Specific hypotheses..... | 51 |
| 2.9 Analysis..... | 53 |
| 3. Results | |
| 3.1 Section Overview..... | 54 |
| 3.2 Overview of data analysis..... | 54 |
| 3.2.1: Initial tabulation and screening..... | 54 |
| 3.2.2 Statistical Analyses and hypotheses testing..... | 55 |
| 3.3 Sample Characteristics..... | 57 |
| 3.4 Participants' engagement with the games and tests..... | 58 |
| 3.5 Guess Who: construct validity..... | 59 |
| 3.5.1 Descriptive statistics and relationship to age and intellectual level..... | 59 |
| 3.5.2 Concurrent validity tests..... | 62 |
| 3.5.3 Fidelity to scoring criteria..... | 64 |
| 3.5.4 Practice effects and reliability analysis..... | 65 |
| 3.5.5 Qualitative description of process..... | 66 |
| 3.5.6 Section Summary..... | 66 |
| 3.6 Connect 4: construct validity..... | 67 |
| 3.6.1 Descriptive statistics and relationship to age and intellectual level..... | 67 |
| 3.6.2 Concurrent Validity Tests..... | 68 |
| 3.6.3 Fidelity to scoring criteria and practice effects..... | 70 |
| 3.6.5 Qualitative description of process..... | 71 |
| 3.6.6 Section Summary..... | 71 |
| 3.7 Performance of ABI group..... | 73 |
| 3.7.1 Overview of Group and performance..... | 73 |
| 3.7.2 Profile of individual ABI participants' performances..... | 76 |

| | |
|---|-----|
| 3.7.2.1: ABI 1..... | 76 |
| 3.7.2.2: ABI 2..... | 77 |
| 3.7.2.3: ABI 3..... | 78 |
| 3.7.2.4: ABI 4..... | 78 |
| 3.7.2.5: ABI 5..... | 79 |
| 3.7.3 Section Summary..... | 79 |
| 3.8 Ecological Validity..... | 80 |
| 3.8.1 Comparison of experimenter’s observations and parental ratings..... | 80 |
| 3.8.2 Ecological validity of GWAS and WI..... | 82 |
| 3.8.3 Section summary..... | 82 |
| 4. Discussion | |
| 4.1. Section Overview..... | 83 |
| 4.2 Summary of main results..... | 83 |
| 4.3 Implications of results..... | 84 |
| 4.3.1 Design of measures: Guess Who..... | 84 |
| 4.3.2 Design of measures: Connect 4..... | 87 |
| 4.3.3 Engagement with games and tests..... | 89 |
| 4.3.4 Ability of games to discriminate ABI and TD group performance..... | 90 |
| 4.3.5 Ecological validity of testing method..... | 90 |
| 4.3.6. Viability of method as a neuropsychological screen..... | 93 |
| 4.3.7 Further study limitations..... | 94 |
| 4.4 General discussion..... | 96 |
| 4.5 Conclusion..... | 98 |
| 5. References | 100 |
| 6. Appendices | 114 |

1. Introduction

1.1 Study Overview

The idea for this study emerged from discussions with clinicians who use Guess Who (©Hasbro Gaming) and Connect 4 (©Hasbro Gaming) informally in the early stages of neuropsychological assessment of children and young people (CYP¹). This study seeks to validate the use of these games as neuropsychological tests and ascertain whether they provide additional benefits to established tests in terms of examinee engagement and ecological validity.

This introduction firstly provides an overview of the main difficulties encountered by CYP following an acquired brain injury (ABI) and the role of neuropsychological testing in their treatment and care. Theoretical and practical issues relevant to the development and potential usefulness of these novel tests are then discussed, including the challenge of engaging CYP with testing, the ecological validity of neuropsychological tests and the adaptation of screening tests for use with younger examinees. The process of adapting Connect 4 and Guess Who as neuropsychological tests, which formed the basis for the hypotheses in this study, is then considered.

1.2 ABI in CYP and the role of clinical psychology

Every year, approximately one out of every 400 CYP in England is admitted to hospital following an ABI (Headway, 2015). The outcomes for the individual, their families and wider society are pronounced. ABI during infancy can significantly affect later academic achievement (Ewing-Cobbs et al., 2006) and traumatic brain injury (TBI) in later childhood is associated with later intellectual impairment and social maladjustment (Cattelani, Lombardi, Brianti & Mazzucchi, 1998). The social and economic effects of ABI in CYP include an increased provision of health and social care, increased levels of unemployment for the injured individual and their families, and an increased likelihood of later physical and mental health problems and imprisonment (Langlois, Rutland-Brown & Wald, 2006).

¹ CYP will be used interchangeably to refer to “children or young people” or “child or young person”

It is widely known that cortical reorganisation following ABI facilitates the partial recovery of function (Bates et al., 2001). However, the response of the brain to early injury is complex. Studies have supported the idea of “early vulnerability”, where CYP affected by an ABI in early childhood (rather than later childhood) subsequently “grow into” cognitive difficulties when more complex skills are expected to emerge (Anderson & Moore, 1995). Several studies point to the existence of a “double hazard” where the severity of injury can interact with early injury age, social disadvantage and lack of stimulation to reduce functional outcomes (see Anderson, Spencer-Smith & Wood, 2011, for review). This highlights the importance of the role of clinical psychologists in identifying deficits following ABI and creating relevant psychosocial interventions, such as supporting enhanced family functioning (Casey, Ludwig & McCormick, 1986). Neuropsychological testing, along with clinical interviewing, behavioural observations and the assessment of mood, is an important tool for clinical psychologists engaged in this work.

1.3 Issues with neuropsychological testing with CYP

This section outlines some theoretical and practical issues relating to neuropsychological testing of CYP with ABI. These issues underpinned test development and were also issues that the study sought to address.

The theory and practice of clinical neuropsychology is grounded in parallel developments in medicine, neuroscience and psychology from the early to mid twentieth century (Lezak, 2012, p. 3-4). Firstly, the medical study and diagnosis of behavioural syndromes following brain injury in adults (prompted by the need to assess wounded servicemen) provided a picture of the localised nature of brain function. Psychologists would later begin to study such individuals to validate and inform their models of human cognition, which were emerging from the work of linguistic theorists such as Chomsky (1959). Alongside this, psychologists such as Binet, Spearman and Wechsler were developing statistically-based tests to capture variances in mental capacities such as intelligence. Clinical neuropsychology is, in effect, an amalgam of these developments. The approach involves establishing the variance across the population in performance on tests measuring specific cognitive functions, many

of which are localised. From this, the effect of brain injury on different cognitive functions for an individual can be estimated. A poor performance on a test can help infer whether a specific brain structure has been compromised by a suspected injury.

Caution is required when translating the conceptual framework of clinical neuropsychology to the study of CYP with ABI (Reed & Warner-Rogers, 2008, pp.1-5). The study of brain and behaviour relationships in CYP takes place in the context of a developing brain. Compared with adults, the effects of ABI on the developmental trajectory of abilities must therefore be considered. For example, dyscalculia can develop due to the presence of visuo-spatial problems in CYP, but the nature of this difficulty is markedly different to dyscalculia following focal ABI in adults (Ansari & Karmiloff-Smith, 2002). Neuropsychological testing with CYP thus raises particular theoretical and practical issues, including:

- i. There is less evidence for localised “modules” underpinning discrete cognitive functions earlier in development. Therefore, the ability being tested must be usually present at the examinee’s age and likely to be affected by ABI.
- ii. As the aetiology of brain lesions in CYP is different to that found in adults (Headway, 2015), the focus in testing will be different if we assume a relationship between pathology, injury site and behavioural, cognitive and emotional sequelae.
- iii. The ecological validity of neuropsychological tests may be questionable, affecting the ability of a clinician to form hypotheses regarding “real-world” difficulties encountered by the examinee (Olson, Jacobson & Van Oot, 2013).
- iv. The validity of test results is compromised without full engagement with the test (Lezak et al., 2012, pp. 153-155), which may be a particular issue when working with CYP.

The following sections provide a broad overview of the latter four issues in relation to this study.

1.3.1 Developmental Considerations affecting test design

When creating a neuropsychological test for CYP, evidence must exist that the underlying construct/ability being measured is present at the examinee's current developmental stage. Otherwise, test data will reflect ability in another cognitive area or be uninterpretable. The impact of developmental level on test selection and interpretation can be illustrated by considering executive functions.

It is well-known that damage to the prefrontal cortex in adulthood can result in impairment in executive functions. However, when testing CYP, we must consider how these skills are developing throughout childhood and adolescence. Working memory and inhibition are known to be present in preschool children (Epsy, Kaufman & McDiarmid, 1999), but skills such as set-shifting and planning ability emerge at approximately eight years and are still not at adult levels by 13 years of age (Davidson, Amso, Anderson & Diamond, 2006) and 12 years of age (Welsh, Pennington & Groisser, 1991). Physiological and functional changes in prefrontal regions are evident throughout adolescence (see Blakemore & Choudhury, 2006, for review), suggesting further development of executive functioning during this period. One can also infer how the trajectory of development of one executive subcomponent influences another; it is clearly difficult to create complex verbal plans in the absence of working memory capacity to keep such plans "in mind".

Consideration of developmental level is therefore vital in designing an appropriate test of executive function. A complicated test of planning and strategic behaviour is unlikely to help build an understanding of difficulties faced by younger CYP. Conversely, executive function difficulties such as planning may not be picked up if only working memory is tested with an adolescent.

1.3.2 ABI aetiology and screening test design

The aetiology of brain lesions varies greatly across age groups (Headway, 2015; Cancer Research, 2015). This implies a need to select different tests for use with different age groups, e.g. to assess the more general impairments due to TBI which are common in CYP (Kolb & Whishaw, 2003, pp. 702-704). This is relevant to the current study as it impacts upon the content of neuropsychological screening tests. The value of screening tests lies in limiting unnecessary time spent on delivering extensive test batteries to relatively unimpaired individuals.

However, as these tests have been created for detecting impairments in dementia (Folstein, Folstein & McHugh, 1975) and mild cognitive impairment (Nasreddine et al., 2005) their extension for use with CYP (Ouvrier, Goldsmith, Ouvrier & Williams, 1993), is problematic. A systematic review of the literature (see Appendix A) shows that eight studies have attempted to address this issue and create screening tests compatible with CYP.

Six papers examined the use of existing tests as screening tools. Lewandowski (1984), Charvet et al. (2014) and Reitan and Wolfson (2004) used single tests of processing speed and attentional skills to screen CYP with various forms of ABI. Lewandowski found that the symbol-digit test had good sensitivity and specificity for impairments in multiple sclerosis. However, both other groups of researchers found that approximately 25% of typically developing children were miscategorised using both the symbol-digit test and Trail-making Test B. Furthermore, as a screen these brief paper-and-pen tests may provide relatively little process information to inform further formulation. Three further papers used broader batteries as screening tests. Two of these (White et al., 2006; Pejnovic et al. 2012) did not provide data on the test's ability to identify deficits following ABI. Krull et al. (2008) found that a small screening test battery (digit span, trail-making, grooved pegboard and verbal fluency) was highly predictive of performance on an extensive battery in a large group of CYP who survived cancer. The presence of three tests that tap executive functions is a strength of this study when considering the aetiology of ABI in CYP. However, digit span may not have uncovered executive function difficulties such as planning in the older participants. Also, the abstract nature of the tests may make it difficult to identify functional difficulties from observations of the examinee's performance, i.e. the tests are likely to have poor ecological validity.

Two groups of researchers designed screening tools for ABI in CYP similar in format to those used with adults. Leiby, Pollock, Mouanoutoua & Lewey's (2014) screening tool consisted of brief subtests in multiple domains, with overall score and subtest scores significantly different between the ABI and control groups. However, the paucity of executive subtests is notable given the prevalence of frontal injury in CYP. Billiard et al.'s (2002₁) screening tool showed good accuracy in distinguishing a large group of children with epilepsy from typically developing

children (Billiard et al., 2002₂). Unfortunately, data for the screen is only available for children aged between four and eight, so its usefulness with older children and adolescents is unknown.

In summary, the development of neuropsychological screening tests appropriate for use with CYP with ABI is at an early stage. The ability of subtests adapted from larger test batteries to identify deficits due to ABI is questionable and this method provides less process information for initial formulation. Screening tests specifically designed for CYP show promise, but issues exist with the focus of the subtests and their usefulness across a wider age range. Also, none of the tests reviewed attempted to address the issue of engagement at the acute stage of injury or the issue of ecological validity.

1.3.3 Ecological Validity of executive function tests

Many definitions of ecological validity have been given (see Gioia & Isquith, for overview), but all emphasise the ability of a test to make useful predictions regarding the examinee's real-world, functional strengths and difficulties. Many neuropsychological tests for use with CYP have poor ecological validity (Gioia & Isquith, 2004). Intervening variables, such as environmental distracters, are absent from testing which leads to an over-estimation of the CYP's real ability. It can also be difficult to decipher "better" or "worse" real-world behaviour to link with a test score (Silver, 2000). This is clearly an issue for rehabilitation work, where a goal is to understand functional deficits and compensate for them by adopting internal and external compensatory strategies.

The issue of ecological validity is of particular relevance to the assessment of executive function, the target of the novel measures in the current study. While executive function is generally related to "supervisory control", the degree of overlap/separation of the executive function subcomponents, and thus their construct validity, is debated (see Rabbit, 1997, for overview). There is even evidence that almost all variability in executive function is accounted for by differences in fluid intelligence (Roca et al., 2010). For a clinician, this means that scores on one test of executive function (e.g. planning) may be highly correlated with deficits on another (e.g. problem-solving) and may relate to executive functions not easily tested (e.g. emotional regulation). This, coupled with the

evidence that the absence of intervening variables on tests leads to ability being over-estimated (see Gioia & Isquith, 2004 for discussion), limits a clinician's ability to predict functional difficulties from executive function test outcomes, whether this is test scores or process observations.

Ecological validity can be raised by increasing a test's *verisimilitude*, roughly the test's face validity in relation to a real-world behaviour that relies on the ability being tested. Attempts to achieve this have broadly followed three strategies (from Fawcett, Payne & Howell, 2007):

- i. Make paper-and-pen tests more like a real-life task.
- ii. Observe and rate real-world tasks in a natural environment.
- iii. Use proxy reports, where parents rate everyday functioning based on their own observations.

In their systematic review of the literature, Chevignard et al. (2012) identified 17 ecologically valid measures for use with CYP that were each developed using one of the above three strategies. One further test has been reported since this review (Gilboa et al., 2015), which uses a virtual reality paradigm to test sustained attention. Chevignard et al. note the strong construct validity of the published observed real-world tasks and the large normative sample data available for the proxy reports. Furthermore, an increased sensitivity in detecting executive function deficits has been reported for observed real-world tasks (Toussaint-Thorin et al., 2013) and for modified paper-and-pen tasks (Longaud-Vales et al., 2016). Rating scales may identify slightly different executive function difficulties compared with traditional tests (see Isquith, Roth & Gioia, 2013 for review) and observed real-world tasks facilitate the identification of where task performance breaks down (Berg, Edwards & King, 2012).

Despite this, issues exist with these tests. Modified paper-and-pen tasks, such as the Zoo Map Test (Wilson et al., 1996) could be viewed as standard planning tests with an added metaphor; it is not clear how verisimilitude is increased. Also, Chevignard et al. note that many such tests are quite structured and omit important intervening variables. Observed real-world tasks (e.g. cooking) may rely on culturally specific skills. The ability of those unfamiliar with the tasks may thus be underestimated. They also assume that children complete tasks in

isolation from others, which may not be typical for children from some cultural backgrounds. Parental rating scales may fail to show a developmental trend (Roy et al. 2015) or simply reflect parental anxiety (Chevignard et al., 2012). Lastly, interaction with the clinician is relatively controlled in all of these tests which could limit key emotional and behavioural information available from the testing process.

Considering executive functions once again, Gioia and Isquith (2004) suggest that, given the demand on executive functions in novel situations, the insertion of lesser degrees of examiner-controlled structure and allowing intervening variables to intervene in the testing process could help determine how deficits emerge in real-world situations. In this study, an attempt was made to introduce these variables via the “gamification” of an executive function test. Game playing could introduce intervening variables such as managing frustration while the cooperation and turn-taking required with the playing partner could provide insight into interpersonal functioning. This data could be gathered via observations during a game-based assessment and provide a useful adjunct to the quantitative data gathered from the results of the game itself.

1.3.4 Engagement during testing

Neuropsychological testing with CYP places a particular onus to engage examinees with the test material. The examinee must be assisted to perform as well as possible if the test score is to accurately represent their ability in the area being tested, rather than reflecting the influence of external or internal contingencies (Lezak et al., 2012, pp. 153). It is easy to discern the impact of external factors on performance. We have a limited capacity for processing incoming information (Lavie, 1995) and external distracters are likely to use up this capacity. This reduces the ability to closely attend to all relevant information on perceptually demanding tasks such as neuropsychological tests.

The impact of intrapersonal variables can be more subtle. Simple vigilance to a task affects performance and this ability, as measured by sustained attention, improves throughout childhood (Lin, Hsiao & Chen, 1999). This implies that a less engaging task could result in poorer performance among younger children regardless of their actual ability in the area under examination. Anxiety is known

to affect performance on neuropsychological tests, e.g. on timed tests of attention and the encoding stage of memory tests (Airaksinen, Larsson & Forsell, 2005). The standard conditions of neuropsychological testing, which can involve adherence to test manual instructions and a lack of reassurance-giving, often in a hospital setting, can be very anxiety-provoking for examinees (Lezak et al., 2012, pp. 154). Younger examinees may be confused about the exact relevance of test results to their care, which could increase anxiety even further. These potential triggers for anxiety should be considered in the context of the examinee's developmental level, as CYP are inherently less likely to be able to regulate such emotions without external assistance (see Music, 2016 for review).

For test creators, the impact of engagement on test outcome therefore means a responsibility to ensure that the test process appears relevant, intuitive and even enjoyable for the examinee.

1.3.5 Section summary

The preceding section reviewed four issues relevant to the development of a board game for use as a brief neuropsychological test.

Firstly, tests should probe for abilities known to be present at the examinee's age, which is particularly relevant to the assessment of executive function. Secondly, test design must be cognisant of the aetiology of ABI within the age group being examined (i.e. TBI in CYP). Existing neuropsychological screening tests developed for use with CYP do not fully address this issue. Thirdly, the issue of the ecological validity of neuropsychological tests used with CYP is only beginning to be addressed. The gamification of tests is one potential way of addressing this issue. Finally, engagement with testing is key to obtaining a valid test score, and this is likely to be a particular issue for CYP who may find the test procedure confusing and anxiety provoking.

The next section outlines in more detail how the gamification of a neuropsychological test may help address these issues and how two board games, Guess Who and Connect 4, might be adapted for this purpose.

1.4 Study aims: can board games be used as neuropsychological tests?

1.4.1 Cognitive assessment via games or play.

This study seeks to establish whether using board games as neuropsychological tests with CYP could help address the issues outlined in section 1.3. A board game with embedded measures could be potentially highly engaging and anxiety reducing. This could be particularly useful for screening at the acute stage of ABI where motivation may be low and anxiety is likely to be heightened. Game playing may also be a daily activity for CYP in a wide range of cultures when compared with some of the ecologically valid tasks already created.

Despite the difficulties defining what constitutes play, there is good evidence of play being a preoccupation of CYP regardless of cultural background (Cohen, 2006, pp. 1-13 for overview), one which has a key role in development, e.g. allowing children understand the inner worlds of others and coordinate their actions accordingly (ibid, p. 57-84). There is a surfeit of play-based assessments in other areas of developmental psychology (e.g. Power & Radcliffe, 2000; Irwin, 2000). However, a systematic review of the literature (see Appendix B) indicates that the potential for direct neuropsychological assessment of older children or adolescents using structured games has not yet been investigated. Some established tests have been adapted to allow CYP engage with testing via familiar technology (i.e., using tablet computers for WISC-IV assessment), but no measure has been created in a bottom-up fashion from play-based activities.

The following section outlines how Guess Who and Connect 4 might be adapted as neuropsychological tests, with particular attention to the potential properties of the embedded measures.

1.4.2 Guess Who and Connect 4 as neuropsychological tests

Guess Who is a two-player game where both players are confronted with a rack of 24 faces with varying attributes (i.e. hair colour, gender, earrings, etc., see). Each player draws one face from a pack and it is their opponent's task to find out which face they have drawn by asking yes/no questions. Connect 4 is also a two-player game, similar to "noughts and crosses". Taking every second turn, each

player places tokens in a rack. The winner is the first player to create a line of four tokens either vertically, horizontally, or diagonally.

To function as neuropsychological tests, Guess Who and Connect 4 must show valid psychometric and neuropsychological properties. Psychometric properties refer to the how performance on the test should be variable across the population, potentially mapping onto a statistical distribution such as the normal distribution. This variability should correlate with other variables known to generally correlate with ability, such as an IQ score or chronological age. For adequate neuropsychological properties, performance on the test should correlate with performance on a test that has already shown construct validity for the area being assessed; e.g. a new test of executive function should correlate with an existing validated test of executive function. This is referred to as concurrent validity (Brooks et al., 2010).

Given the large number of neuropsychological tests already published, a new test should provide additional benefit to the psychologist and examinee rather than simply replicating the properties of an existing test. The following sections discuss how aspects of performance on Guess Who and Connect 4 might map onto existing neuropsychological test measures and provide additional benefit by addressing issues discussed in Section 1.3.

1.4.3 Potential psychometric and neuropsychological properties

The format of Guess Who is very similar to that of the parlour game “Twenty Questions”. Tests modelled on Twenty Questions distinguish adults with frontal lobe lesions from typically developed individuals (Baldo et al., 2004), showing that this task may probe executive functions. Lezak et al. (2012, pp. 628-629) outlines how this task relates to frontal/executive function. Firstly, a higher level concept must be formed from particular instances (i.e. asking “Is it a form of transport?” to decipher whether a car, bicycle and submarine are potentially the answer). Secondly, a strategy must be followed that eliminates the maximum number of alternatives on each question.

Norms for Twenty Questions (Delis, Kaplan & Kramer, 2001) show that a developmental progression is evident in task performance. Identifying a verbal

higher-level concept from particular instances is measured as part of the WISC-IV test of general intelligence (Wechsler, 2004), which indicates that Twenty Questions should also tap general intellectual ability. However, one difference between Guess Who and Delis, Kaplan & Kramer's (2010) Twenty Questions is that the shared features across faces in Guess Who are not necessarily abstract. By way of example: identifying that many faces have earrings requires less abstract concept formation when compared to identifying that many items are "cutlery". Thus, one might expect Guess Who to test strategy formation more than concept formation. The need to identify common features and eliminate faces with these features where necessary could also test visual search and attention skills, probed by commonly used-tests such as the Trail-Making Test Part A (TMT Part A, Lezak et al., 2012, pp.422).

Connect 4 requires the examinee to attend to their own performance and that of their opponent and shift between "attacking" and "defensive" play where appropriate. Dividing attention and "shifting set" in this manner is probed by tests of executive function, such as the Modified Card Sorting Test (Nelson, 1976, see Lezak et al. 2012, pp. 636 for review of evidence) and tests of divided attention, such as the TMT Part B (Reitan, 1958, see Lezak et al., 2012, p.423 for review). Both the TMT Part B (Yochim et al., 2007) and the MCST (Grafman, Jones & Salazar, 1990) are sensitive to frontal executive dysfunction in adults. Associations have also been demonstrated between the MCST and IQ scores (Strauss, Sherman & Spreen, 2006, pp. 526-545 for review) and between the TMT Part B and the processing speed aspect of IQ (Sanchez-Cubillo et al., 2009). Use of the TMT with CYP is discussed further in Section 2.5.3.

Finally, both Guess Who and Connect 4 require players to continually monitor their performance and strategy. For example, in Guess Who a slightly riskier strategy may be more advantageous if one's opponent is nearing a correct answer. With Connect 4, there may be several options open to a player on a given move and the risks and benefits associated with each option must be thought through. Monitoring performance in this manner is probed in tests of planning, such as the Zoo Map Test (Wilson, Alderman, Burgess & Emslie, 1996), where examinees must consider options before proceeding to act. Wilson et al. report a moderate negative correlation between competent Zoo Map

performance after ABI and reports of executive function difficulties on a proxy report of everyday functioning.

1.4.4 Observational data and ecological validity

In a standard neuropsychological assessment, observational data can be obtained from observing the examinee's behaviour during interview and assessment. How markedly this behaviour deviates from the "norm" can help inform the formulation derived from test data, answers given on interview and collateral information (Lezak et al., 2012, pp. 164). However, as test conditions are generally constrained and formal, the level of problematic behaviour that arises during testing may be minimised. Testing via game-playing may address this. For example, it may demonstrate whether the examinee shows evidence of competitiveness, of over-familiarity, an ability to take turns, whether they become inappropriately excited about winning, disgruntled or angry about losing or show little joy or motivation in playing the game.

These process observations relate to the potential ecological validity of this method of testing. A challenge of development is learning to coordinate one's interaction with other CYP in a non-egocentric manner, which depends on emotional regulation skills (Carr, 2006, pp. 21-25). The emergence of this ability is known to link with the development of frontal lobes (Gerhardt, 2015, pp. 32-56) and is also dependent on anterior temporal lobe functioning (Glosser et al. 2000). These areas are known to be most affected by TBI (Levin, Culhane & Mendelsohn, 1993), which, as already discussed, is the most common type of brain injury in CYP. It is therefore unsurprising that changes in personality (Ylvisaker, Jacobs & Feeney, 2003) and difficulties with emotional regulation (Eslinger, Biddle & Grattan, 1997) are prevalent in CYP with TBI.

Positioning the neuropsychologist alongside the CYP in a two-player game is a marked deviation from the standard process of neuropsychological testing. This method may lead to more issues with interpersonal behavioural and emotional regulation emerging during testing. How it feels for family or friends to interact with the CYP could become more apparent and assist with neuropsychological formulation. The importance of this is highlighted by the complexity of the origin of behavioural difficulties in CYP with TBI; it is known that CYP with inflicted TBI

show greater emotional regulation problems than those with accidental TBI (Ewing-Cobbs et al. 1998) and that premorbid behavioural problems are highly prevalent in TBI (Tate, 1998). This extra dimension of the testing process could increase ecological validity by providing the examiner with insight into concrete behavioural difficulties. This, coupled with the test results, could feed into a clinician's formulation of executive function difficulties.

1.4.5 Guess Who and Connect 4 as screening tests

As already discussed, Guess Who and Connect 4 can potentially function as tests of executive function and visual attention. Section 2 will show how Guess Who can also function as a test of visual recall memory. By using both games, three domains of cognitive functioning can therefore be briefly examined; executive functioning, attention and memory. This breadth of domains fulfils one requirement of an appropriate screen. Furthermore, since frontal lobe injury is prevalent in CYP with ABI, this focus on executive function deficits makes the games more likely to pick up on acquired difficulties than screening tools with less of a focus in this area.

1.5 Research Questions

The following study examined the use of Guess Who and Connect 4 as a neuropsychological screening tool in a group of CYP aged between eight and fourteen years. Healthy participants and participants with acquired brain injury of varying aetiology were included in the study.

The research questions were

- i. Do novel measures of executive function, visual attention and visual memory derived from Guess Who and Connect 4 show validity as neuropsychological tests?
- ii. Will the performance of healthy CYP on these measures differentiate them from CYP with ABI?
- iii. Does this gamification method provide valuable information regarding the functional deficits and strengths of the examinees? I.e., is the method ecologically valid?

2. Methods

2.1 Epistemological framework and methodology

The most prominent methodological aspects of neuropsychology might appear to place this study within the epistemological framework of scientific realism. The lesion-deficit model provides a foundation for the methods of investigation of neuropsychology and the interpretation of behaviour, particularly after ABI (for extensive discussion, including successes of the model, see Lezak, 2012 p. 101-116). This model assumes that species-wide cognitive functions exist, which are underpinned by discrete, localised neural structures that are similar across individuals.

Some empirical challenges to the lesion-deficit approach include the evidence for cortical reorganisation following brain injury (Bates et al., 2001), the importance of wider connectivity within the brain to support cognitive functions (Friston, 2011) and the reliance of certain cognitive functions on different brain areas at different points in development (Blakemore & Choudhury, 2006). Considering CYP in particular, the lesion-deficit model is challenged by the difficulty differentiating higher-level cognitive functions early in development (Karmiloff-Smith, 1992). This relates to a challenge to the model at a more epistemological level. It has been argued that certain supposedly discrete cognitive functions, executive functions in particular, cannot be conceptually separated (Rabbitt, 1997, pp. 1-18). By this argument the terms *disinhibition*, *perseveration* and the inability to *shift set* could be applied to the same behaviour depending on the context within that behaviour is played out.

One reaction to this is to adopt a qualitative approach to understanding brain injury, which might focus on an individual's unique phenomenological experiences following brain injury (see, e.g. Sacks, 2015) rather than assessing an individual's ability relative to normative scores from the wider population. However, a practical concern is that this approach is less relevant for establishing a method of efficiently screening a large number of CYP for cognitive impairment.

The critical realist stance outlined by Bhaskar (1997), provides an epistemological framework that can accommodate the successes of

neuropsychological methodology but also its empirical shortcomings and conceptual critiques. In this framework we can separate relatively intransitive knowledge (that brain functioning and ABI affect behaviour) from transitive knowledge (that interpretation of this behaviour, and thus epistemology, is affected by context). Thus, critical realism supports a materialist ontology that underpins scientific investigation (and thus quantitative methodology) but invites a critical stance in how such results are interpreted. In this quantitative psychometric study of tests and measurement, we will accept that brain structure and ABI affects observed behaviour, but also that the constructs being measured and the means of measurement are imperfect. Therefore, it will not be assumed that all CYP performing poorly on tests of set-shifting will each have a frontal lobe lesion and have similar difficulties moving between tasks. However, this data will be seen as potentially informing a contextualised interpretation of an individual's observed difficulties.

2.2 Design

The study used a correlational design to establish the concurrent validity of the novel neuropsychological measures with the established measures. All typically developing (TD) participants completed both novel and established measures. Statistical analyses were used to establish whether performance on the novel measures predicted performance on the established measures as per the hypotheses. As discussed in Section 3.2.2, establishing concurrent validity is an important step in demonstrating that a novel measure actually measures the construct in question.

Unfortunately, given the low number of ABI participants a between-subjects comparison of ABI and TD participants was not possible. Considering ecological validity, a statistical analysis of the examiner's ability to categorise strengths/weaknesses in the group based on observations was also not possible due to low participant numbers. Therefore, a case study methodology was used to investigate both of these areas.

For the ABI participants, individual case studies attempted to identify whether the performance of individual ABI participants on the novel measures fell outside the "normal range" of the TD group. Data from the established measures was also

examined to ascertain if they also placed the ABI participants' performances outside the range of the TD group and (where data was available) the normal range of the population as a whole.

For ecological validity, the examiner's ratings of difficulty and strengths in behavioural regulation and metacognition were compared with those made by parents (for ABI's and TD children) on the standardised proxy report of executive function difficulties. This was to demonstrate whether those rated as having strengths/weaknesses in either of the aforementioned domains by parents would also have been identified as having such strengths or difficulties by the examiner.

2.3 Participants: Selection, recruitment and power analysis

Participants with ABI (N=5) were recruited from a neurorehabilitation service in the South of England. TD participants (N=14) were initially recruited from after-school clubs linked with schools and children's centres within London and the Home Counties. Snowball sampling was used as participants recruited from these sources distributed the information sheets via email and social media to acquaintances. Additionally, some siblings of participants with ABI expressed an interest in participating having read the information sheets and were recruited. All participants opted to complete the study tasks at home. Demographic information on participants is provided in Section 3.3 of the results.

The following inclusion criteria were used for all participants:

- Aged between 8 and 14 years at the time of testing. As outlined in the introduction, this age range is where many executive functions begin to emerge and develop.
- Native English speaker or schooled entirely via English from point of admission to primary school. This criterion sought to reduce the influence of language skills as a confounding factor (particularly on Guess Who performance).
- An absence of pre-morbid learning difficulties or pervasive developmental disorder. As the study sought to measure deficits in specific cognitive abilities, this criterion sought to exclude the influence of more general intellectual difficulties.

Inclusion criteria applied specifically to participants with ABI were:

- Acquired brain injury (acquired at any point after birth) due to neurological disease or TBI. As already discussed, the age at which injury occurred affects the developmental trajectory of abilities. However, to maximise recruitment ABI at any age was considered.
- An absence of severe sensory impairments, severe aphasia, severe memory impairments or pervasive attention difficulties (including severe working memory difficulties). All of these issues were likely to impede upon engagement with the tasks, which were intended to measure less severe deficits than these.
- An absence of severe mood or anxiety difficulties. Along with making participation in research distressing for the participant and thus unethical, the effect of anxiety on attention and therefore test performance would have been a confounding factor.

One TD participant aged 7.10 years who was particularly keen to participate was included. To reduce respondent burden, he did not complete the demanding Zoo Map Test.

A staff member at the recruitment site identified approximately 60 potential participants with ABI on a patient database that were likely to meet the inclusion criteria. The staff member posted information sheets and contact details of the study coordinator (the author) and supervisor to these potential participants. This ensured that patient information was not shared without prior consent and that participants initiated contact with the study coordinator.

Using G Power 3.1 (Faul, Erdfelder, Lang & Buckner, 2007) the anticipated number of participants required to reach adequate statistical power ($\alpha=0.05$, $\beta=0.8$) for the correlational design was calculated as 23. This assumed a moderate correlation of $r = 0.5$ between the validated measures of executive function and the novel measures. The latter was a relatively conservative estimate given the strong correlation of executive function tests with one another (Roca et al., 2010). As the number of participants recruited did not reach this 23, a post-hoc analysis of power will be provided with all correlations relating to the tests of concurrent validity.

2.4 Guess Who and Connect Four: test development

2.4.1 General description of games

Guess Who is a two player game based around a set of 24 cartoon characters. The objective is to identify a target character present within the other player's set of characters either by eliminating all other alternatives or by guessing the target character's name at any point in the game.

At the start of a game, each player's set of characters are presented on a specially designed board (see Figure 2.1). Each character is standing vertically up and facing the player so they cannot be seen by the player's opponent. The target character for the other player is picked at random from a deck. The two players alternate asking yes-no questions about the target card to eliminate characters (e.g. "is the character male?", "are they wearing a hat?"). The board is designed in such a manner as to allow eliminated characters to be turned down (see Figure 2.1).

Both players view the exact same set of faces placed in different positions on their respective boards. Only the face of each character is depicted with the first name printed beneath the face.

Connect Four is a two-player game. The objective is to be the first person to form a line of four tokens in a row in a standing rack (see Figure 2.2), either vertically, horizontally, or diagonally. The rack contains 42 positions arranged linearly in a rectangular grid of six rows and seven columns. Tokens are dropped from the top of the grid and occupy the lowest available position within the column they are dropped into.

Players take alternate turns placing their chosen colour tokens (red or yellow) in a vertical standing rack. Connect 4 thus involves both constructing a line and simultaneously stopping one's opponent from constructing a line.

2.4.2 How could these games be used as tests?

The motivation for the current study emerged from discussions with clinicians who use Guess Who and Connect Four in their work with CYP with ABI. They use the two games to provide observational data on executive function and



Figure 2.1: A Guess Who board during a game with five characters eliminated



Figure 2.2: A completed Connect 4 game won with four red tokens arranged diagonally.

attentional difficulties before conducting detailed neuropsychological assessments. An investigation of the value of this process would require an understanding of the cognitive skills required to successfully complete the games. Therefore, the author spent time becoming very familiar with tests by playing them extensively with adult peers and, for Connect 4, also with online algorithms (Maths Is Fun, 2014). A Google search was used to look for solutions to the games, i.e. strategies guaranteed to win the games. This period of research provided an understanding of how the process of game-play related to the process of commonly used neuropsychological measures. It also gave an indication of how benchmarks or poor, acceptable and proficient performance might be set. The proceeding sections (2.4.3 and 2.4.4) detail the outcome of this work, showing how a formalised method of “administering” and scoring Guess Who and Connect 4 was arrived at.

2.4.3 Test Development: Guess Who

2.4.3.1 Creating a character set

To make Guess Who a viable test, a question or set of questions asked by the examinee must show relative advantage in identifying the target quickly and accurately. Considering this, it was noted that in a single set of Guess Who faces the majority of traits were only held in common by between one and three faces. This meant little difference between the number of alternatives eliminated by different questions. For example, with only three females, three characters with blue eyes and three characters with curly hair, there is no relative advantage to asking whether the target character has either of these traits. The same number of characters is likely to be eliminated in each case. To resolve this issue, three different versions of Guess Who were scrutinised and a composite set of faces was created from two of these versions (see Appendix C). Traits were distributed to a greater or lesser extent within this set (see Appendix D). This ensured that the range of questions that players could ask would be more-or-less advantageous. The proceeding section details the rationale for this.

2.4.3.2 Game-play strategies

Guess Who is similar in format to the parlour game “Twenty Questions”. This game requires one player to decipher the object or concept that another player has in mind using a series of yes/no questions. Neuropsychological tests have already been created in this format (e.g. Delis, Kaplan & Kramer, 2001).

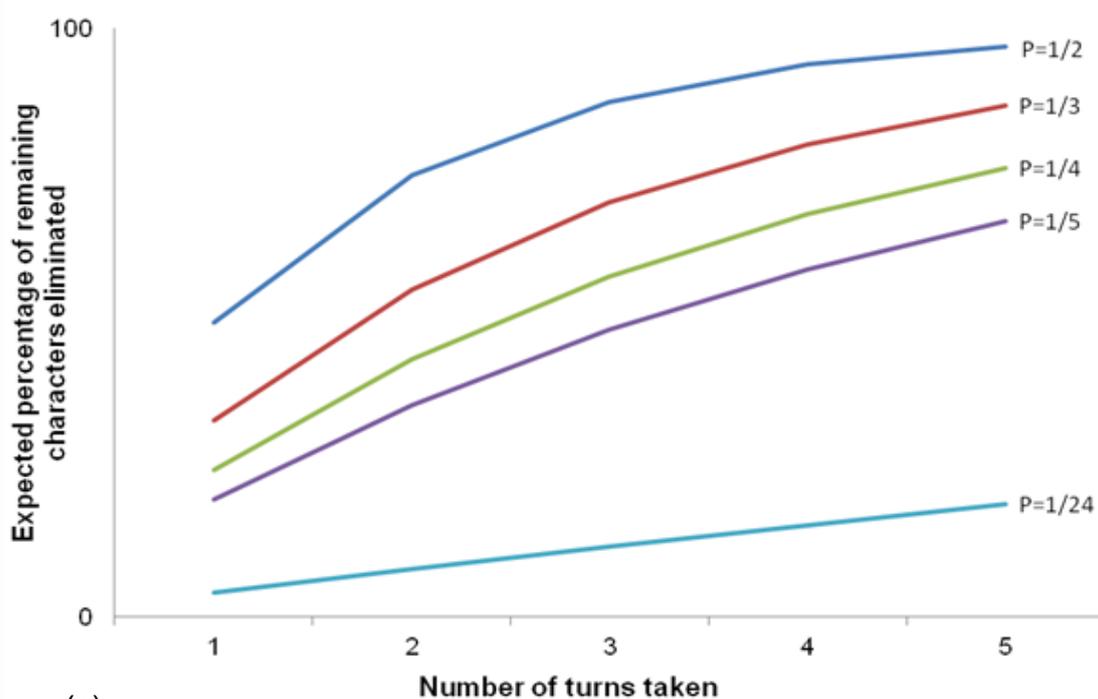
From preliminary investigations, a common sense strategy for any Twenty Questions game appeared to involve asking questions guaranteed to eliminate half of the alternatives on each turn. This strategy is termed *lose-half* here. For example, in Guess Who if half of the remaining characters have glasses, and only a quarter have brown hair, the most advantageous strategy is to ask “Does the character have glasses?” This will guarantee the elimination of 50 percent of the remaining characters. Asking whether the character has brown hair carries a one-in-four chance of eliminating 75% of characters, but is more risky given the three-in-four chance of eliminating 25%.

One might question whether the riskier strategy is indeed less advantageous than *lose-half* in the long run. Considering the previous hypothetical example, a player could continue to ask questions regarding attributes shared by only a quarter of the remaining characters. This player is odds-on to eliminate 75 percent of the characters at least once in every four turns. Perhaps this strategy will in fact get to the correct answer just as quickly as *lose-half*?

It can be demonstrated that this does not happen. If we define P as the minimum number of alternative answers eliminated by each question (e.g. this will be $\frac{1}{2}$ for *lose-half* and $\frac{1}{4}$ for the alternative strategy described above). Also, if we define E_n as the minimum number of characters eliminated on turn T_n and n as the number of the turn (1st, 2nd, 3rd, etc.). The expected percentage of characters eliminated on any turn for any strategy is then:

$$E_n = 100 * (E_{n-1} * (P * (1 - E_{n-1})))$$

The output from this equation can be plotted for all values of P , as shown in Figure 2.3 (a). This demonstrates when an increasingly riskier strategy is selected, an increasing number of turns are required to eliminate a similar number of characters.



(a)



(b)

Figure 2.3: Graph (a) demonstrates that lose-half ($P=1/2$) eliminates characters faster than any riskier strategy when the most probable answer to each question is given direction. $P=1/24$ shows the outcome for guessing individual names in Guess Who.

Graph (b) depicts the number of games won by strategies of varying risk over a large number of games (Rober, 2015). In the long run, less games are won by riskier strategies.

This still does not fully account for when the riskier strategy “gets lucky”. For this, the above equation can be used to model the outcome of all Twenty Questions – type games (Rober, 2015). So, we consider every combination of answers for

every level of strategy, i.e. we consider games on which the riskier strategy is lucky on one, two, three etc. questions or lucky on every question. Using this approach, it is still seen that lose-half is more successful than all riskier strategies on the majority of games it played (Rober, 2015). The resulting distribution of wins is normal, with lose-half at the mean/median point of the curve (see Figure 2.3 (b) for illustration), with increasingly riskier strategies further to the edge of the curve.

Now to consider how this could relate to cognitive skills. Lezak et al. (2012, p.628) argue that 20 Questions brings to light aspects of executive functioning, particularly concept formation. For Guess Who this would mean forming an idea of what a target character looks like, i.e. "I am searching for a bald man". Lezak et al. identify different questions relevant to the purpose of assessing conceptual formation. This includes constraint seeking questions, which narrow down alternatives ("is he smiling?") and pseudo-constraint seeking questions, which are irrelevant and don't narrow down alternatives ("Does the dog have paws?", "Is he a friendly fellow?"). Clearly, the former indicates intact concept formation whereas the latter indicates difficulties in this regard.

However, in addition to concept formation, it is clear from the previous discussion of lose-half that forming a successful strategy is vital in identifying the target character as quickly as possible. Observing a complex concept within the characters (e.g. "Does he look malevolent?") will have no advantage if it eliminates few characters from consideration. Delis et al. (2001) note this in their version of Twenty Questions. They describe a "spatial strategy" (i.e. "is the target on the left of the board?") that relies little on concept formation but arrives at the target quickly. This demonstrates the independence of concept and strategy formation in Guess Who and suggests that both should be measured independently. Therefore, two separate measures of executive function were created. "Guess Who Strategy Score", which is analogous to Delis et al.'s Weighted Achievement Score, is a measure of quality of strategy and planning ability (posited components of executive function, Lezak et al., 2012, pp. 671-683, Bechara, 2007). The number of pseudo-constraint seeking questions, analogous to Delis et al.'s Set-Loss Questions measure (Delis et al., 2001, p. 159), is taken as an indicator of difficulties with concept formation.

Outside of the central task, Guess Who can be used to gather data on visual skills and memory. Eliminating characters requires intact visual search and attention skills. Thus, performance on established tests of this skill, such as the Trail-Making Test Part A (Reitan, 1958), could be analogous to “Elimination Errors” made on Guess Who.

Considering memory function, Guess Who contains discrete visual information (faces) placed within a predictable grid. This has some topographical similarity with abstract figure recall tasks. These test the ability of examinees to recall the spatial location of visual details relative to a set grid. For example, in the Rey Complex Figure (Rey, 1941), the BMIPB Figure Recall (Coughlan, Oddy & Crawford, 2007) and the RBANS Figure Recall (Randolph, 1998) successful recall of the material requires the examinee to construct a large rectangle with 2 diagonals around which further visual details are arranged. Asking examinees to recall where on the board faces were located after playing Guess Who may tap the same abilities as figure recall tasks.

A summary of the novel measures derived from Guess Who, the constructs they are hypothesised to measure, and the established measures of the same constructs are presented in Table 2.1 below.

Table 2.1

Summary of novel measures derived from Guess Who and established measures hypothesised to measure the same construct.

| <i>Construct</i> | <i>Games Measure</i> | <i>Validated measure</i> |
|-------------------------------|---|--|
| Concept formation | Guess Who pseudo-constraint-seeking questions | D-KEFS 20 Questions set-loss questions |
| Strategy Formation | Guess Who Strategy Score | D-KEFS 20 Questions Weighted Achievement Score |
| Visual search and cancelation | Guess Who Elimination Errors | TMT A time TMT A errors |
| Visual Recall | Guess Who face location recall | RBANS Figure recall score |
| Planning ability | Guess Who Strategy Score | Zoo-Map Errors |

2.4.3.3 Novel Guess Who measures: scoring procedure

- Guess Who Strategy Score (GWSS): This examines the ability to identify, implement and monitor an optimum strategy (i.e. lose-half) and make decisions while new information becomes available during the game. As already described, this is considered an aspect of intact executive functioning. For each turn, the lowest number of potential characters eliminated by the question was identified (i.e. for either a yes or no answer) and divided by the number of remaining characters. Thus, 50% was the highest possible score. This was averaged across turns in the 5 games for a final score.
- Planning: As strategy formation and planning ability are conceptually very similar (Lezak et al., 2012, pp. 671-683), GWSS was also identified as a measure of planning.
- Pseudo-constraint seeking questions: A measure of concept formation. Questions irrelevant to the purpose of narrowing down alternatives, including bizarre, tangential, inherently subjective and repeated questions (including those that are inherently repeated questions, i.e. “Is he angry?” after being told the character is sad) were counted here. The score given was the number of pseudo-constraint seeking questions divided by the total number of questions.
- Elimination errors (EE): The total number of faces erroneously eliminated or not eliminated from consideration for the entire five games played.
- Guess Who visual recall: For the immediate recall score, directly following completion of Guess Who the participant was sequentially presented with 6 faces that were not target cards. With all characters in the rack turned down, the participant was asked to point to where on the board they saw each face. A correct response scored four and one point was awarded for selection of a location immediately adjacent to the target. After a 20 minute delay, a standard delay to ensure that recall engages long-term memory (e.g. Baddley et al., 1994) the procedure was repeated to collect the delayed recall score.

2.4.4 Test Development: Connect Four

2.4.4.1 Establishing what constitutes a “good” performance

Early investigations involved playing Connect 4 many times over with an online algorithm (Maths Is Fun, 2014). The algorithm allowed one to vary the computer’s

game-play from “very easy” (near-random token placement) through to “hard” (computer using a “solved” solution to the games, see Section 2.4.4.3 below). Heuristics for game-play were also obtained via Google searches. This helped to derive an understanding of what constituted a good performance, which in turn informed how participants’ performances would be assessed.

2.4.4.2 Preliminary definitions

The following terms are used during this and subsequent sections:

Ahead and Behind: A player is ahead if they are closer to lining up four tokens than their opponent, who, conversely is behind. For example, Player A is ahead of Player B if (s)he only needs two tokens to complete a line of four while player B requires three tokens to complete a line of four.

Blocking: actively moving to stop opponent from continuing a line of tokens.

Building: actively continuing to create a line of tokens.

Algorithm: a sequence of step-by-step instructions that can be coded as a computer programme.

Heuristic: Intuitively recruiting a solution for a similar, simpler problem to a problem of greater complexity (Kahnemann, 2011, pp. 98).

2.4.4.3 A heuristic for Connect Four

From a mathematical perspective, Connect Four is a “solved” game. An algorithm called “minimax” is mathematically proven to win every game it starts. Even if minimax “plays itself” the algorithm taking the first turn is guaranteed to win on the final turn (see Higginbotham, 2012, for overview). So, like Guess Who, for Connect Four an optimal strategy exists to compare with a player’s performance.

However, there are important differences between the two games that affect how we might compare performance to the respective optimal strategy. In Guess Who, lose-half is a relatively intuitive example of “narrowing-down”. By contrast, the minimax algorithm is highly complex for those unfamiliar with computer science (see Higginbotham, 2012). Also, in Guess Who, even the performance of random guessing can be easily compared with lose-half by the examiner. In

Connect Four, the optimal strategies are complex and prescriptive rather than probabilistic and appear impractical to implement in a neuropsychological test scoring grid.

With this in mind, a heuristic was instead identified that greatly increases the likelihood of winning a game of Connect 4 compared to random moves.

Kahnemann (2011) provides extensive evidence that heuristics are used by people when confronted with novel tasks of the complexity of Connect Four. A simple heuristic that is successful in tic-tac-toe games such as Connect Four is (MIT, 2010):

1. Complete a winning move where available.
2. Block an opponent's winning move.
3. Maximise opportunities by playing towards the centre of the board.

Playing Connect Four while developing the measure, it became apparent that these steps are inadequate in guiding choices with no winning moves available. Steps one and two were expanded to account for this. Also, step three was omitted as it appeared intuitive only following game-playing experience.

Therefore, the final heuristic was:

1. Complete a winning move where available.
2. Block an opponent's winning move.
3. Build a line unless continuing to do so will result in opponent winning.
4. Block opponent's line unless allowing opponent to continue will result in you winning.

This heuristic, which I have called *switch-search*, is presented in Figure 2.4. Each participant's performance was assessed by comparing their choice on each turn with the "choice" the heuristic would make. Two components of *switch-search*, *win-search* and *switching*, were considered specifically relevant to cognitive skills and these are now described.

Win-search relates to vigilance for a winning move becoming available on every turn. One's opponent can place a token in one of 7 positions on each turn, opening up numerous possibilities that might not have been previously recognised, including winning moves. Therefore, on each turn it is important to

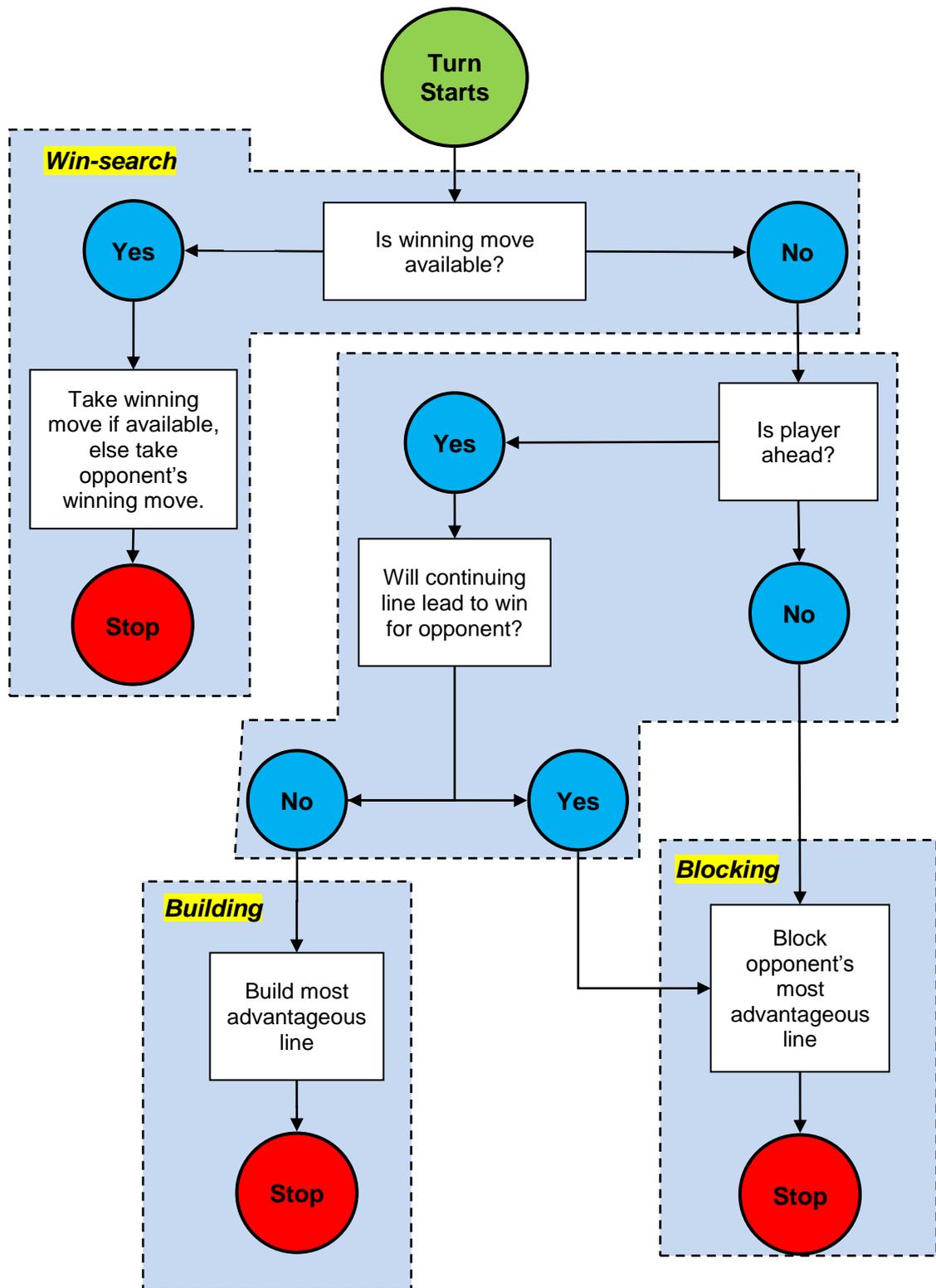


Figure 2.4: Switch-search heuristic with components

search the board for a winning move for either player. Psychologically, win-search could therefore tap skills of visual search and attention.

Switching relates to how, when there are no winning moves available, one player is generally ahead and building a line while the other player is behind and blocking. A player blocking will be unable to win if they continue to play defensively when the opportunity to go ahead emerges. Similarly, a player building will lose if they continue to build when their opponent moves ahead of them. Psychologically, moving between these two components can be related to divided attention and set-shifting. Players have to divide their attention between their own play and their opponent's to determine who is ahead. When this changes, a player must shift from blocking to building or vice versa.

An obvious example of switching is at the start of the game, where the player going first is automatically one token ahead and building. Their opponent will need to switch to building if they are to win. Figure 2.5 illustrates the importance of switching a little later in a game. In (a) the player using red tokens has been ahead for the entire game. By continuing to build lines they have failed to switch to blocking their opponent's line and winning move. Conversely, in (b) red has been blocking their opponent. Their last defensive move stopped them from moving ahead and putting yellow in a defensive position. These examples demonstrate that "switching" from blocking to building or vice versa is key to advantageous game-play.

2.4.4.4 Novel Connect 4 measures: scoring procedure

- Wins identified (WI): Related to win-search, this is the total number of winning moves taken or blocked during games as a percentage of total winning moves available. Where two winning moves were available on a turn to a player or their opponent, this was counted as one winning move. This is hypothesised to be a measure of visual search and attention.
- Switching score (SS): The total number of switches correctly made by a player during a game as a percentage of switches available. Where a switch was not made immediately but correctly made on a subsequent move, this was counted as one correct switch (i.e. 1/1). An erroneous switch (one made where it should not have been taken) was counted as a missed switch.

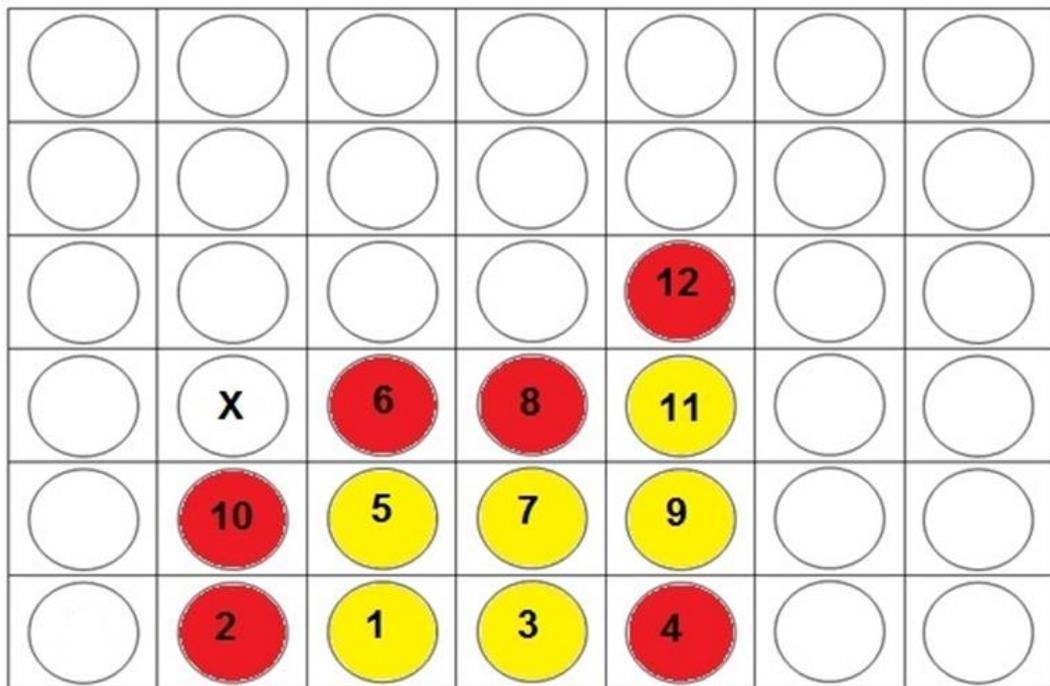
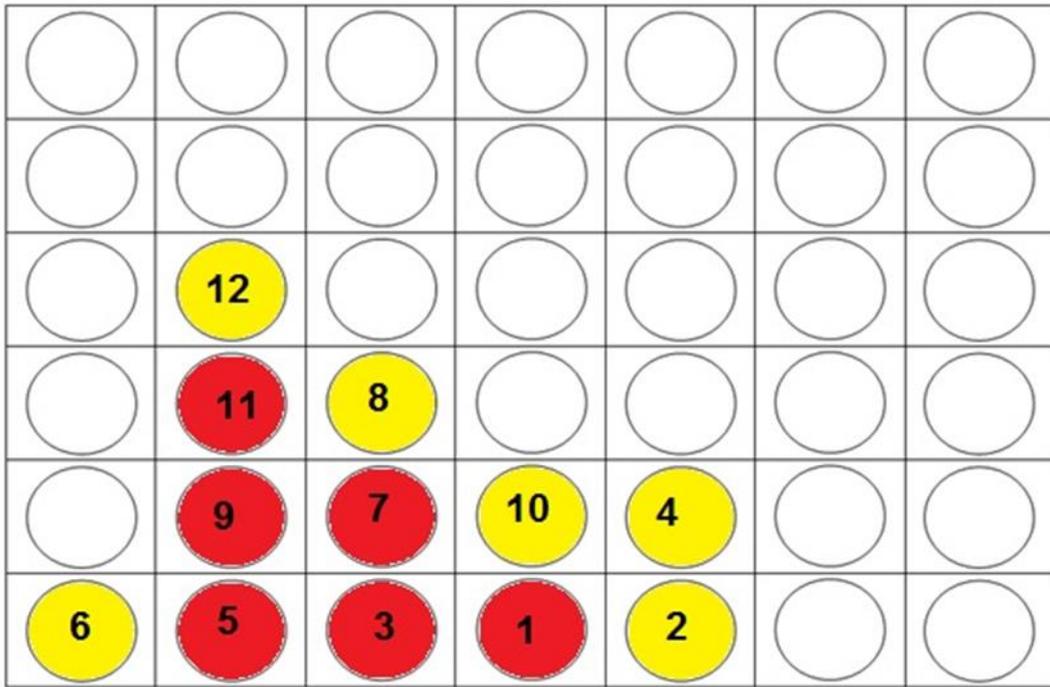


Figure 2.5: Failure to switch. In the top figure, red has started the game building but failed to switch to blocking, allowing yellow to win at move 11. In the bottom game, red could have built a line at X, but, having been blocking for most of the game, has failed to set-shift and placed a token at 12. Numbers denote the move number (1st, 2nd, 3rd, etc.)

Table 2.1

Summary of novel measures derived from Connect 4 and established measures hypothesised to measure the same construct.

| <i>Construct</i> | <i>Games Measure</i> | <i>Validated measure</i> |
|-------------------|------------------------------|----------------------------|
| Visual Search | Connect Four Wins Identified | TMT A Time TMT A errors |
| Set-shifting | Connect Four Switches | MCST errors |
| Divided attention | Connect Four Switches | TMT B –A |

2.5 Neuropsychological battery details and test instructions

2.5.1 Respondent burden

Respondent burden was explicitly considered when designing the study. This was particularly important for participants with ABI, who may have experienced greater stress in participation, for example due to fatigue. All participants were provided the option to complete the study at home if desired (which all participants did). The shortest, least-burdensome tests were also included in the battery where possible. Thus, including a break (at least 10 minutes and up to 20 minutes if desired) the study took between 80 and 120 minutes for the participants to complete.

2.5.2 Games and novel measures

Participants completed the games/novel measures first and in the following order.

Pre-test screening

Although participants were pre-screened, a further screening of gross cognitive difficulties was undertaken before testing (see Appendix E). This consisted of easy one item tests of visual scanning, verbal expression, verbal comprehension, auditory attention and verbal working memory. These are commonly included in validated screens of cognitive impairment (e.g. Folstein, Folstein & McHugh, 1975; Mioshi et al., 2006; Nasreddine et al., 2005).

Guess Who

The variables measured are described in Section 2.4.3.3.

Five games were played in total to ensure that adequate data would be gathered. The participant was asked how familiar they were with the game before beginning to ascertain whether they had regularly played the game or had never played it before. They were then asked to select a colour rack that they preferred (red or blue). The participants went first on each game. The participant and experimenter selected target faces for each other from a deck before each game. These were not returned to the deck for subsequent games (participants were made aware of this). Detailed verbal instructions and prompts, which were created to closely mirror those of D-KEFS Twenty Questions (Delis, Kaplan & Kramer, 2001) are given in Appendix F.

Responses were recorded using the response sheet shown in Appendix G. The experimenter recorded each question asked by the participant. Using the grid shown, the characters eliminated by the participant was recorded following each question by placing a “1” on the grid after question one, a “2” following question two, etc. The experimenter’s first three questions were set as per Appendix H. After the third question, the experimenter asked questions that attempted to prolong the game as much as possible (i.e. using riskier questions where the participant was slower at arriving at the target). Following the fifth game the immediate recall trial was conducted, which was repeated after 20 minutes (See Appendix F for instructions). The locations selected by the participant for each face were recorded in the response sheet in Appendix I.

Qualitative information seen as relevant to the assessment of cognitive difficulties was recorded during Guess Who (and Connect 4). This included issues such as missed turns and the participant’s emotional reactions and was informed by the subscales of the BRIEF (Gioia et al., 2000).

Connect Four

The variables measured from Connect Four are described in Section 2.4.4.4.

Five games were played in total to ensure that adequate data would be gathered. The experimenter went first in each game. The participant was asked how familiar they were with the game before beginning to ascertain whether they had regularly played the game or had never played it before. They were then asked to select a colour token they preferred to play with (red or yellow). Detailed verbal instructions and prompts provided to the participant are given in Appendix J.

In the first, third and five games, the experimenter began by placing a token in the central column. Once blocked, the experimenter started a new line from a single token, which provided the participant with an opportunity to begin attacking. The experimenter placed tokens close to the centre of the board to maximise opportunities. The experimenter opened opportunities when ahead for the participant in an attempt to provoke switches. When behind, the experimenter actively attempted to block the participant and get ahead, in an attempt to force the participant to switch to blocking.

On the second and fourth games, the procedure was slightly modified. Here, the experimenter placed the first token in either the left or rightmost column and continued to play to the edges of the board in general. This was to maximise opportunities for the participant. A similar strategy was followed otherwise.

Each move was recorded on the response sheet (see Appendix K). A number was written into the space on the board occupied by each move, which allowed the game to be analysed after testing by colouring alternate moves red and yellow in numerical order. Qualitative information seen as relevant to the assessment of cognitive difficulties was recorded during the games. This included issues such as missed turns and the participant's emotional reaction to playing the games with the experimenter.

Participants counted the number of red and yellow tokens after each game without pointing, to further rule out any issues in lower-level perception that may have affected outcomes.

2.5.3 Battery of established measures

Figure 2.6 depicts the psychometric, clinical and practical issues that were considered when compiling the battery of established measures. Note that this

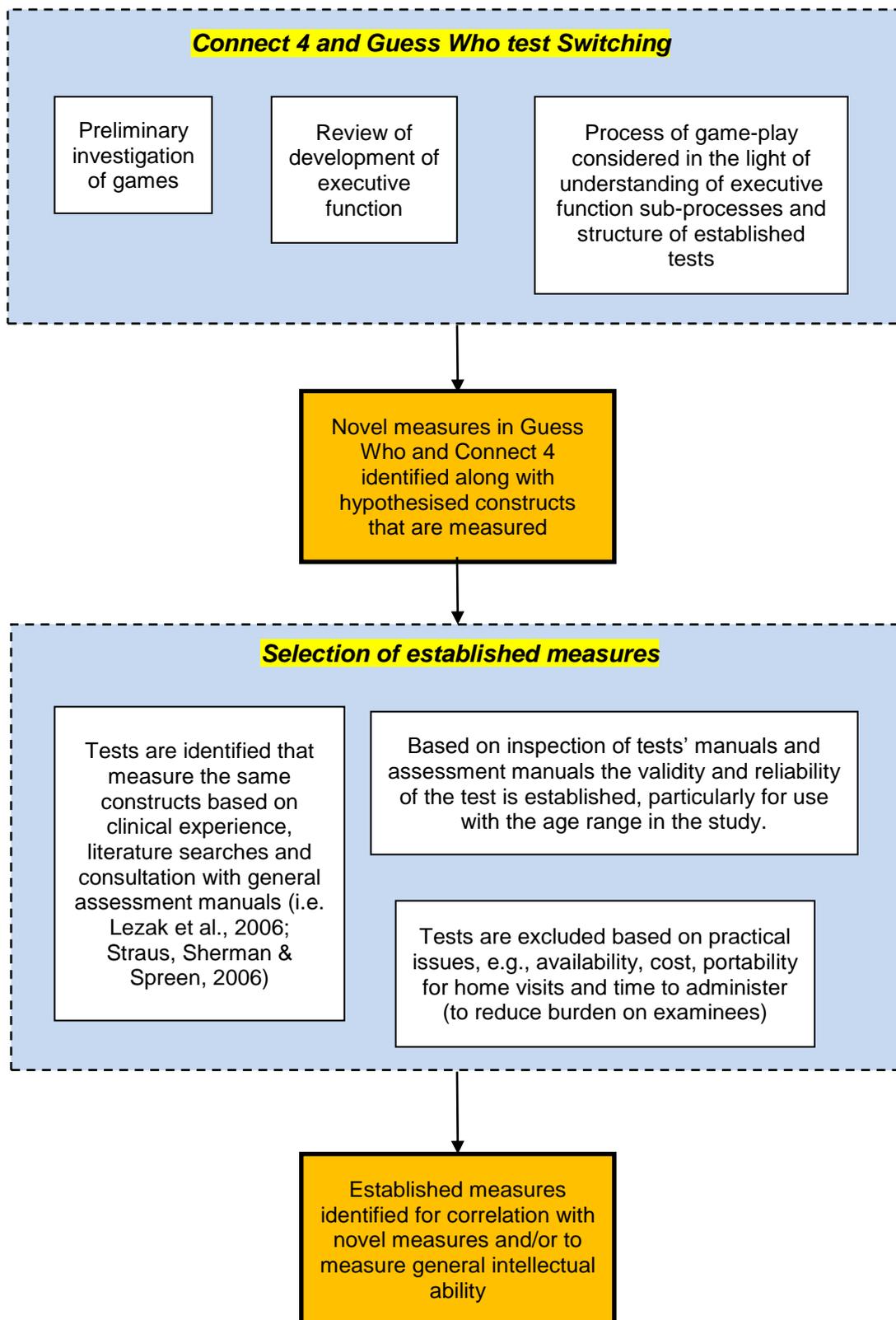


Figure 2.6: Decision process for selection of tests for battery of novel and established measures.

process was not linear and some issues applied more to certain tests than others. The participants completed the battery of established measures after the games/novel measures and following a break in the proceeding order. Instructions were taken from test manuals unless otherwise stated.

RBANS Figure Recall (Randolph, 1998)

This figure recall task is usually administered in a copy, immediate recall and delayed recall format. Points are scored for the number of components of the figure correctly copied and recalled. Correct sizing, orientation and placement of the figure components underpins scoring.

The copy trial was administered as per the author's manual. The recall task was modified to reduce the load of free recall and thus make it more topographically similar to the Guess Who memory task. Participants were presented with the main large rectangle with two diagonal lines (see Appendix L) that forms the outline of the figure and told:

"This is an outline of the previous figure you copied. I am going to present you the remaining parts of the figure and I would like you to draw them where you saw them". If the participant commented on the orientation of the components, the experimenter said: "The parts may not be exactly in the same direction as you remembered them on the figure, but I would like you to place them facing in the correct direction."

They were the shown the following eight missing components (See Appendix L):

- Short arrow
- Three small circles
- Large Circle
- Large Box
- Double curved narrow "S" shape
- Large Cross
- Single Horizontal Line
- Two linked lines forming two vertices of a triangle.

The components were presented in their original size but not necessarily in their original orientation to the participants. Similar to the original administration method, points were awarded for correct placement, orientation and sizing of the components (see Appendix M). Blacked out items in Appendix M indicate where no score was awarded as the items were considered too easy or redundant.

Twenty Questions (Delis, Kaplan & Kramer, 2001)

In this test, the examinee is presented with a stimulus page with 30 pictures of common objects (See Figure 2.7). The examinee is asked to identify an unknown target on the page in the fewest number of yes/no questions. The objects can be subsumed within various subcategories. For example, the stimulus pictures include 15 non-living things, within which are 7 items of machinery, within which are four vehicles, within which are two vehicles that fly. The most effective problem-solving strategy is to eliminate half of the objects on each turn regardless of whether the examiner answers yes or no. The examinee has 20 questions to find the target, at which point the test is discontinued.

The authors state that executive functions tapped by this test include the ability to perceive the various categories and subcategories and the ability to incorporate the examiner's feedback to formulate an efficient questioning strategy. Data for CYP that demonstrates a developmental progression in performance are



Figure 2.7: D-KEFS Twenty Questions, section of stimulus sheet

provided in the authors' manual. Adult patients with frontal lobe damage have been shown to require more questions to arrive at an answer and make more simple guesses ("is it a dog?") than patients with damage to other brain areas (Upton & Thompson, 1999).

Performance was assessed via:

- i. Weighted Achievement Score. This closely parallels the participant's fidelity to a strategy of eliminating half of the remaining alternatives on each question.
- ii. Counting the number of set-loss questions (tangential questions, conceptually incoherent questions or those that do not allow a yes-no answer).

Trail Making Test (Reitan, 1958)

The Trail Making Test (TMT, Reitan, 1958) is a widely used test of processing speed, divided attention and executive functioning (Lezak, 2012, p.422). The test consists of two parts (see Figure 2.8). In Part A (TMT A), following a short practice trial the examinee is presented with an A4 page on which the numbers 1 to 25 (digits, circled) are distributed pseudo-randomly. The examinee is asked to draw a line between each number in order without lifting the pen from the page. Part B (TMT B) is similar except on this occasion the numbers one to thirteen and letters A to L appear on the page. The examinee is asked to link the numbers and letters in order, while alternating between numbers and letters (i.e. A, 1, 2, B, etc.).

Time of completion of each part is recorded as are errors. The time for TMT A is subtracted from that of TMT B to eliminate the effect of speed on TMT B performance. Performance on TMT A correlates with that of other timed visual search tasks in adults (Sanchez-Cubillo et al., 2009). Performance on TMT B also taps speed and attention skills in adults along with executive function as measured by the Modified Card Sorting Task (Kortte et al., 2002). The performance of CYP in the TMT shows an improvement with development, both in terms of speed and accuracy (Anderson et al., 1997, cited in Strauss, Sherman

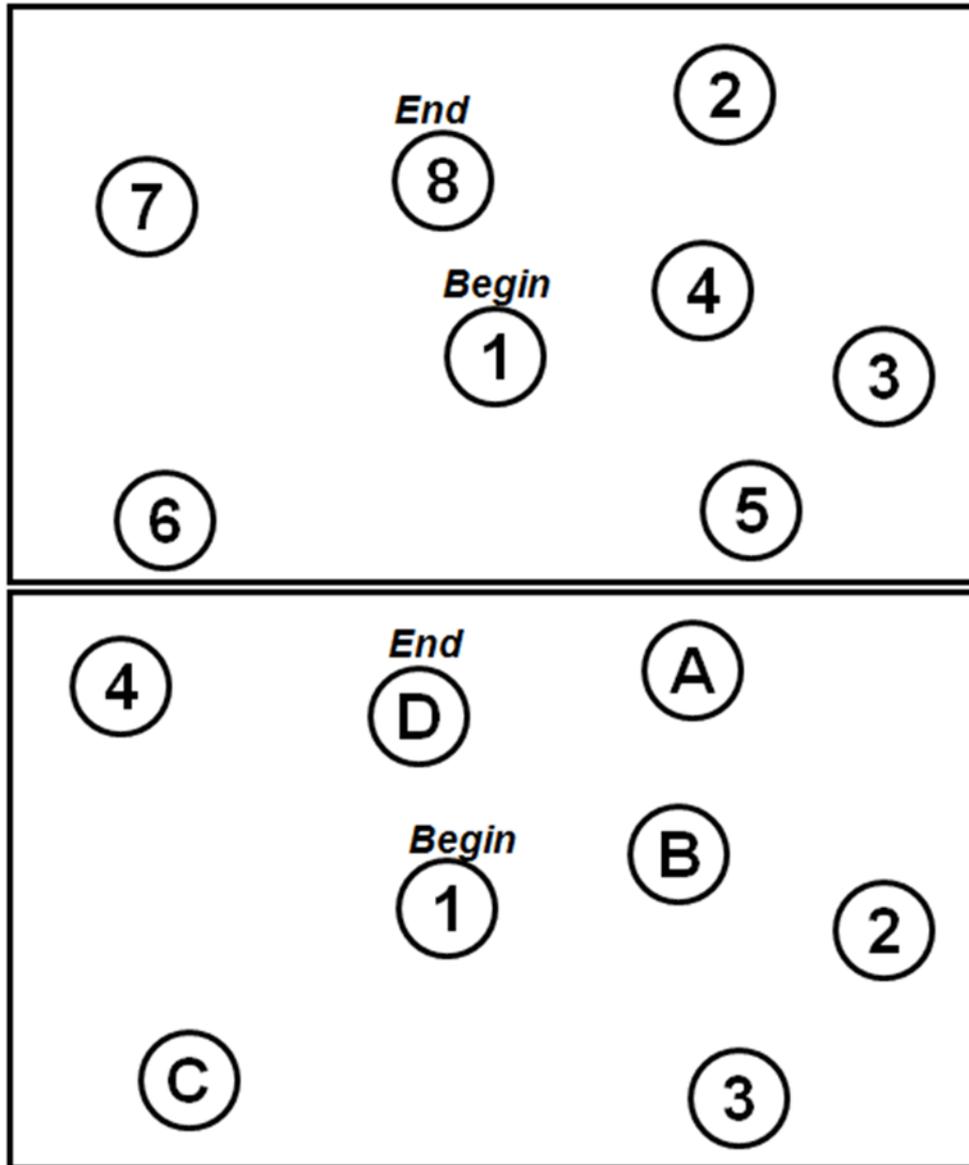


Figure 2.8: Trail Making Test, Part A practice trial (top) and Part B practice trail (bottom)

& Spreen, 2006, p.). CYP with ABI have been shown to perform worse than age-matched controls on a shortened form of the test (Reitan, 1971)

Scoring and instructions were based on those provided by Strauss, Sherman & Spreen (2006, p.655-677). Time to complete and number of errors was recorded. The original test format (Reitan, 1958) was used.

Modified Card Sorting Test (MCST, Nelson, 1976)

This test was used as a measure of set shifting and perseverative errors.

The administration method of Nelson (1976) was adopted. Four “key cards” are placed in a horizontal line on front of the examinee (See Figure 2.9). The key cards are distinguished from each other by the number of objects (one, two, three and four) the shape of objects (squares, circles, triangles and crosses) and the colour of the objects (red, blue, green, yellow) that appear on them. There are 48 additional cards in a pack to sort alongside the key cards according to one of three sorting rules (shape, colour, number). Nelson found that adult frontal lobe patients made more perseverative errors and sorted fewer correct categories than typically developed adults. Normative data for the MCST is available for children, where a developmental progression is apparent in performance (Cianchetti et al., 2007).

Number of correct categories and errors were recorded. Errors were perseverative responses (repeating last response when feedback had just been given to indicate that this response was incorrect) unexpected errors (where the participant left the current category unprompted) and bizarre responses (where it was unclear how the card was related to the target card, i.e. three red crosses matched to two green squares).

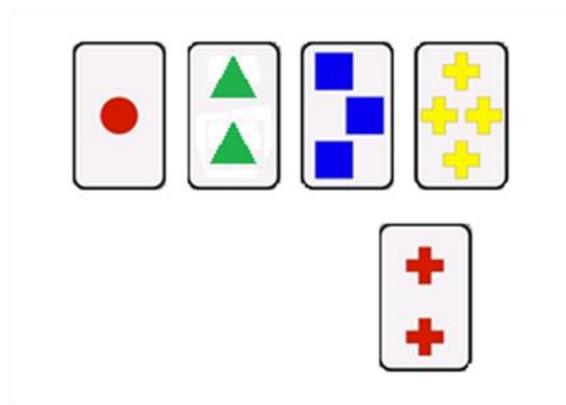


Figure 2.9: Modified Card Sorting Task, showing four key cards (top) and a single card to match to the key card (bottom).

Zoo Map Test (Wilson et al. 1996)

This is a test of planning from the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C). It is a paper-and-pen task designed to be ecologically valid by simulating a real-life situation of planning.

Examinees are given a map of a zoo (see Figure 2.10) along with instructions for the task. These detail:

- i. Locations to visit.
- ii. Rules to follow while making the visit (i.e. places and paths that can only be passed through once or on any number of occasions).

There are two versions of the task. In each, examinees must visit six of the locations on the map. Version One is more taxing as participants must decide the order in which to visit the six locations that leads to a route that minimises rule-breakages. In Version Two, the visiting order is set, which drastically reduces the planning demands.

Examinees are explicitly informed that accuracy is more important than speed when executing the task. Planning skills are measured by calculating the sequence in which locations are visited, number of rule-breakages, planning time (before beginning the task) and execution time for the task. A developmental progression in performance has been demonstrated for the age range considered here (Engel-Yeger, Josman, & Rosenblum, 2009).

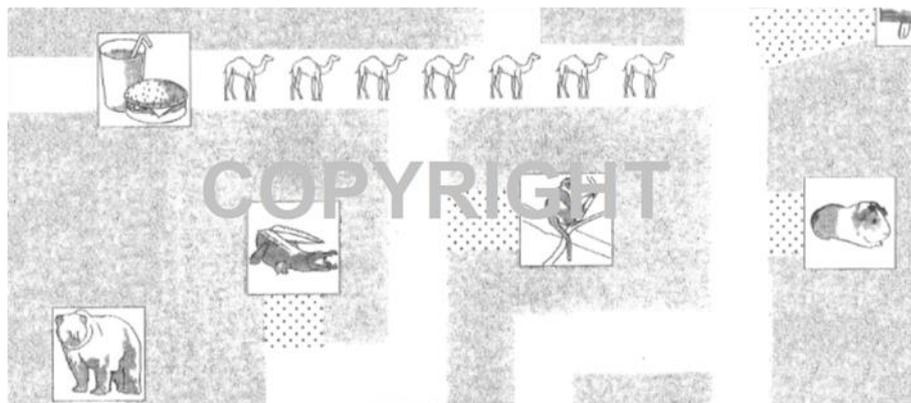


Figure 2.10: Section of map from Zoo Map Test (Versions One and Two)

Raven's Standard Progressive Matrices (Raven, Raven & Court, 2003).

This was used to test non-verbal intellectual ability. It is a test of spatial reasoning known to load highly on the *g* factor of general fluid intelligence (Raven, Raven & Court, 2003). Similar matrix reasoning tests are included in other popular tests of intellectual ability (e.g, the WISC-IV, Weschler, 2003), wherein they correlate strongly ($r=0.9$) with performance IQ subscales (Flanagan & Kaufmann, 2004). Administration of the full matrices set (60 items), was not possible due to time constraints on testing. A truncated nine-item version was instead administered, which has an extremely strong correlation ($r=0.98$) with performance on the full form of the test (Bilker et al., 2012). Overall score was recorded.

WISC-IV Similarities Subtest (Weschler, 2004)

This subtest of the WISC-IV measures verbal abstraction and concept formation and was used as a test of general verbal ability. Pairs of words are presented and examinees are asked how the two words are alike ("How are anger and joy alike?"). Items increase in difficulty as the subtest progresses. Scores of 0, 1 or 2 are given based on increasingly accurate and abstract responses. Similarities scores has a correlation of $r=0.9$ with the Verbal Comprehension Index from the WISC-IV (Flanagan & Kaufmann, 2004), showing that it is a good predictor of verbal ability in general. Overall score was recorded.

Behaviour Rating Inventory of Executive Dysfunction, parent form (BRIEF, Gioia, Isquith, Guy & Kenworthy, 2000).

This is a proxy report of executive function difficulties. It was designed to address issues of ecological validity with standard executive function tests. The parent rates the CYP's behaviour for the last three months. The scale consists of 86 items that describe manifestations of executive function difficulties. Frequency of behaviour is rated on a Likert scale (never, sometimes, and often). Eight sub-scales are obtained: Working Memory, Initiate, Plan/Organize, Organization of Materials, Inhibit, Monitor, Emotional Control & Shift. Overall Metacognition, Behaviour Regulation and Global Executive Indices are derived from these sub-scales. Normative data from 1419 TD children and 852 children from clinical groups is presented in the manual, demonstrating the expected progression with

age and higher incidences of difficulties within the clinical groups. Further discussion of rating scales is presented in Section 1.2.2.

In the current study, a parent of each participant completed the BRIEF. The experimenter noted whether difficulties associated with any of the subscales / indices was apparent from observations during Guess Who and Connect 4. Note that this rating took place in the break in testing, before completion of the standardised tests and before scores on Guess Who and Connect Four were calculated.

Ratings of test enjoyment and anxiety

Two Likert scales were created to assess how enjoyable and anxiety-provoking the participants found both the games and the standardised tests (see Appendix N). Participants were provided with a pen to complete the scales and were explicitly instructed to be as honest as possible in their ratings.

2.6 Ethics

Ethical approval for the study was granted by an NHS Research Ethics Committee (Reference: 16/WM/0331) and the University of East London Research Ethics Committee. Each participant and their parents/guardians were provided with information sheets at least 24 hours before participating in the project (see Appendices O, P). Before beginning the research tasks, participants and their parents/guardians were encouraged to ask questions about the research tasks or request clarification regarding the contents of the information sheets. Consent was sought from each participant in accordance with the Declaration of Helsinki (2013). As each participant was age 14 or under, signed consent for each CYP to participate was obtained from a parent/guardian and each CYP provided written assent (see Appendices Q, R).

Participants were advised that participation could be withdrawn at any point without disadvantage to themselves, including during testing. Verbal and non-verbal signs of anxiety, fatigue and any other signs that may have indicated a preference to discontinue testing were monitored for. All participants were asked at the break if they were happy to continue with the tasks.

As agreed with participants and as per the terms of ethical approval, all participants were assigned a unique, anonymising identification code (i.e. P1, P2 etc.). This code, and not the participant's name, appeared on all data sheets associated with the research including the parental proxy report, which were kept in a locked cupboard. A spreadsheet identifying the participant codes was kept in a separate location on an encrypted, password-protected USB key.

2.7 General procedure

All participants requested to complete the research tasks at home in the evenings after school or at the weekend. Ahead of completing the tasks, participants' parents were informed that a quiet room free of visual and auditory distraction would be required for completion of the research tasks. During testing, the participants sat opposite the researcher at a table. A break of between 10 and 20 minutes (at the participant's choosing) was provided between completion of Guess Who / Connect Four and the neuropsychological tests. The instructions for each test were read slowly and chunked into short phrases to facilitate encoding and comprehension. Particular attention was paid in this regard with participants with ABI where there may have been more subtle issues with working memory. Language from the test manuals was simplified to aid understanding where necessary. Participants were asked to repeat the instructions for each task before beginning. Where comprehension or memory of the instructions was incomplete, they were reiterated until understood and fully recalled.

2 participants (both with ABI) requested that a parent stay in the room while the tasks were completed. In both cases, the parent sat behind the participant to eliminate visual distraction. They were also requested to not provide prompts, feedback or reassurance to the participant. A report on performance was provided where requested once data collection for the entire study was gathered.

2.8 Specific hypotheses

For Guess Who, the following were the primary hypotheses for the tests of concurrent validity. Note that some measures were excluded due to insufficient responses, as detailed in Section 3.2.1. This reduced the number of planned hypotheses.

- Hypothesis one: Performance on Guess Who, as measured by GWSS, shows concurrent validity with an established measure of strategy formation, the D-KEFS 20 Questions Weighted Achievement Score.
- Hypothesis two: Performance on Guess Who, as measured by GWSS, shows concurrent validity with an established measure of planning ability, the Zoo Map Test (Version 1, error score).
- Hypothesis three: Performance on Guess Who, as measured by EE, shows concurrent validity with an established measure of visual search, the TMT Part A (completion time).
- Hypothesis four: Recall of the position of characters on the Guess Who board shows concurrent validity with a measure of recall memory function, the modified RBANS Figure Recall score.

For Connect 4, the following were the primary hypotheses for the tests of concurrent validity:

- Hypothesis one: Performance on Connect 4, as measured by WI, shows concurrent validity with a measure of visual search, the TMT Part A (completion time).
- Hypothesis two: Performance on Connect 4, as measured by SS, shows concurrent validity with an established measure of set-shifting, the MCST (number of errors made).
- Hypothesis three: Performance on Connect 4, as measured by SS, shows concurrent validity with an established measure of divided attention, the TMT Part B-A.

Hypothesis testing involved the TD group alone. This was to ensure that validity could also be assessed by examining the relationship with age and general intellectual ability in a typically developing group. A secondary hypothesis was that none of the novel measures derived from Guess Who or Connect 4 would correlate highly with the other established measures (i.e. that they would show good discriminant validity for the constructs they were intended to measure).

Given the low number of participants, hypothesis testing was not possible for the data derived from the ABI group and for the investigation of ecological validity. A case study design was used here as described in Section 2.2.

2.9 Analysis

Details of the data analysis procedures are provided in Section 3.2.

3. Results

3.1 Section Overview

This section firstly details data screening procedures and decisions made regarding the statistical analyses and validity tests undertaken. A profile of the TD sample is then provided. The results from the measures of engagement and tests of validity are then detailed. The performance of the ABI group is then compared with the performance of the TD group. The final section looks at the parental responses on the proxy report of executive function difficulties and what this reveals about the ecological validity of Guess Who and Connect 4 as neuropsychological tests.

3.2 Overview of data analysis

3.2.1: Initial tabulation and screening

Data were transferred from scoring sheets and tabulated for statistical analysis in SPSS Version 23.0 (IBM Corporation 2015). Data was tested for normality using the Kolmogorov-Smirnov test at a significance value of 0.05. Outlying data was screened for using a criterion of $z = \pm 2.58$, i.e. above the 99th or below the 1st percentile of the normal distribution. Given the low number of participants, outlying data was transformed where possible using a square-root transformation to reduce skew and maximise inclusion of data. Thus, for TMT B - A the square root of the raw score was used in the analysis.

The established neuropsychological tests contained multiple outcome measures. While this might provide useful information for clinical practice, including multiple measures of the same test increases the likelihood of a type I error as statistical comparisons may have been effectively duplicated. Therefore, one measure from each respective test was chosen for the analysis. Correlations were used to assess whether multiple measures appeared to measure the same underlying ability/construct in this sample. Where a correlation of $r \geq \pm 0.5$ was observed, one of the inter-correlating variables was selected for the statistical analysis. Variables were also chosen based on the absence of extremely high and low scores. Therefore, due to the absence of perseverative errors for most of the sample on the MCST, the total number of errors was chosen for the analysis.

Table 3.1

Summary of variables from established tests used in hypothesis testing

| <i>Test</i> | <i>Variable</i> | <i>Construct examined</i> |
|-------------------------|--|------------------------------------|
| Zoo Map Test | Errors | Planning Ability |
| D-KEFS Twenty Questions | Weighted Achievement Score | Strategy Formation |
| Trail Making Test | Part A: Completion Time Part B Time - Part A Time | Visual search Divided attention |
| MCST | Total Errors | Set-shifting |

There was a near-absence of set-loss questions on the Twenty Questions Test (seven participants made an error, with only two participants making more than two errors) and a low number of pseudo-constraint seeking questions on Guess Who (eight participants made errors, all making three or less errors). Given this apparent ceiling effect on these measures of concept formation, a statistical analysis comparing these variables was not made. Similarly only two participants made an error on the TMT (one TD participant, one ABI), so these variables were also not included in the analysis. Table 3.1 shows the variables from the established tests that were used to test the primary hypotheses.

Considering the games, two participants (both typically developing) performed erratically on one game each of Guess Who, which made scoring each game impossible. In each case, the scores from the remaining four games were prorated to provide an equivalent score for five games. Each game of Connect 4 was scored by recreating each game move-by-move using the recording sheets. To ensure fidelity to the scoring criteria and highlight any issues with scoring, the experimenter selected five games each of Guess Who and Connect Four and re-scored the games while blind to the original scores given for the games.

3.2.2 Statistical Analyses and hypotheses testing

All hypotheses involving correlations were testing using a Pearson's Product Moment Correlation test statistic, or where data was not normally distributed a Spearman's Rho test statistic (denoted as r_s). All within-subjects' comparisons were made using a non-parametric test, the Wilcoxon Signed Rank test. The

significance level was set at 0.05 for all tests. Where multiple correlations involving conceptually similar variables were made, a Bonferroni correction was used to avoid increasing the likelihood of a type-I error. Thus, for correlations of Guess Who Strategy Score (GWSS) with the measures of strategy formation and planning ability, the corrected significance level was 0.025. For correlations of Switching Score (SS) with the measures of set-shifting and divided attention, the significance level was also set at 0.025.

Based on Strauss, Sherman & Spreen (2006, p. 19), the following were scrutinised where possible to assess the construct validity of the measures derived from Guess Who and Connect 4 as neuropsychological measures:

- Properties of data to assess viability as psychometric measures.
- Concurrent validity with established measures (the noted hypotheses).
- Discriminant validity with established measures of other domains.
- Validity as a screen, based on ability to distinguish ABI and TD scores.

Concurrent validity testing includes the typically developing (TD) group only. The number of TD participants means that the study is under-powered as per the predicted correlations (see Section 2.3). Therefore, a post-hoc assessment of the power of the correlations is presented using the qualitative descriptors for associations described by Hinkle, Wiersma, & Jurs (2003). Coefficients of determination (R^2) were used to estimate the percentage of variance in one variable explained by the other. Discriminant validity did not form part of the hypotheses, but will be discussed post-hoc by examining the correlations of the Guess Who and Connect 4 measures with established measures of different constructs. Therefore, any conclusions of discriminant validity must be tempered by the lack of correction in significance levels and a danger of “dredging” for significant correlations post-hoc.

Performance on cognitive and neuropsychological tests generally improves throughout childhood and adolescence in line with maturation of the nervous system and the associated improvements in cognitive skills and processing speed. As valid neuropsychological measures will reflect such improvement, the relationship between age and the variables extracted from Guess Who and

Connect 4 was analysed. Also, given the close relationship between executive functions and general intellectual ability (Roca et al., 2010), the relationship between these variables and the measures of verbal and non-verbal general intellectual ability were also analysed.

Given the small sample size, statistical tests that compared the ABI and TD groups and that examined the ecological validity of Guess Who and Connect 4 were not possible. Therefore, a case study approach was used. The ABI participants' performance on the novel and established measures were compared with that of TD participants. Where normative data was available, this was used to ascertain whether ABI participants who scored poorly on the normed, established measures also scored relatively poorly on the novel measures. For the analysis of ecological validity, the examiner's categorisation of participants' strengths and weaknesses were examined. This was achieved by contrasting the examiner's categorisations with those made by parents on the proxy report.

3.3 Sample Characteristics

The final sample consisted of five participants with ABI and 14 TD participants. Descriptive data for the TD group is provided in Table 3.2. The ABI group are discussed further in Section 3.7 below. All participants who provided data (four from ABI group, 12 from TD group) stated that they were right handed. In terms of ethnicity, the entire ABI group were White British. Four of the TD group were from a mixed ethnic background and the rest were White British. All participants

Table 3.2

TD group: demographic data and intellectual ability (N=14 unless stated)

| <i>Age Yrs (SD)</i> | <i>Gender F/M</i> | <i>Age parents completed education Yrs (SD)</i> | <i>Raven's Matrices¹ (SD)</i> | <i>WISC-IV Similarities² (SD)</i> |
|-------------------------|-----------------------|---|--|--|
| 10.67 (2.19) | 9/5 | 19.79 (2.68) | 34.80 (13.67) | 23.64 (6.91) |

¹ N=13. Prorated score shown (range:0-60), derived from raw score on short-form test (range:0-9)

² Raw score gave a mean standard score of 12.25 with a standard deviation of 2.30.

were either native English speakers or had been educated entirely through English.

Performance of the TD group on the WISC-IV Similarities Subtest, which provides a good estimate of general verbal ability, was relatively strong. Although not strictly a normally distributed scalar variable, for the purposes of illustration the age-standardised score of the group was 12.25 (SD=2.30) compared to a national mean of 10. Seven TD participants scored in the average range, four in the above average range and one in the superior range. For the measure of non-verbal ability, the Raven's Standard Progressive Matrices (RPM), commenting on the participants' age-standardised scores is difficult given the quite wide ability range that the prorated score for each participant covers. However, using the 1979 UK norms provided by Raven, Raven & Court, (2004) and deducting 3 points to allow for a Flynn Effect within a western educated sample (Pind et al., 2003), it could be tentatively stated that 11 of the 13 participants who completed the test scored broadly in the average range, with two scoring in the superior range. No errors were noted on the screening tasks completed before testing and after Connect 4.

3.4 Participants' engagement with the games and tests

Subjectively, all of the participants appeared to engage well with both the games and the tests. Also, none of the participants appeared to be overly perturbed by the challenge of completing either the games or tests. In line with these observations, using the Likert scales the participants tended to rate both the games and tests as being fun and arousing little anxiety.

A statistical analysis was used to compare the participants' ratings of how enjoyable the games and the tests were. It was hypothesised that the games would be more enjoyable for the participants. Given the low number in the ABI group, both groups were combined for this analysis. The data were non-normal and therefore a non-parametric repeated-measures test was used. Contrary to the hypothesis, the mean rating for the games (M = 4.25, SD = 0.58) was not significantly higher than the mean rating for the tests (M = 3.92, SD = 0.83) $Z = -1.508$, $p = 0.138$.

Given the floor effect on the participants' ratings of anxiety, a statistical analysis of this data was not possible. However, looking at Table 3.3., it appears that very little anxiety was aroused by either playing the games or completing the tests. In fact, and most relevant to rationale for this project, none of the ABI group reported experiencing any anxiety while playing either the games or the tests.

Half of the TD group experienced "a little" anxiety while completing the tests. However, it must be noted that the tests took somewhat longer to complete than the games and were completed after the games. Boredom or an eagerness to complete the tasks may therefore have had more of an effect on participants' ratings for the tests.

Table 3.3
Participants' ratings of anxiety evoked by tests and games.

| <i>Rated anxiety*</i> | | <i>None (1/4)</i> | <i>A little (2/4)</i> | <i>Quite a bit (3/4)</i> |
|-----------------------|-----------------|-----------------------|---------------------------|------------------------------|
| Games | TD Group (N=12) | 11 | 2 | 1 |
| | ABI Group (N=5) | 5 | 0 | 0 |
| Tests | TD Group (N=12) | 7 | 7 | 0 |
| | ABI Group (N=5) | 5 | 0 | 0 |

*Figures indicate the number from each group who chose each rating

3.5 Guess Who: construct validity

3.5.1 Descriptive statistics and relationship to age and intellectual level

Table 3.4 provides descriptive statistics for each of the variables derived from Guess Who that were used in the analysis. Analysing these statistical properties can provide insight into the suitability of these variables as psychometric tools. For example, whether numerous cut-offs along a normal curve might exist to delineate qualitatively different performances. GWSS shows a very low standard deviation, roughly 4.5%, meaning that a large number of participants scored close to the mean. Despite this, inspection of the participants' raw data shows that scores across games for individual participants varied greatly, by up to fifteen

Table 3.4

Guess Who: Descriptive statistics for TD group (N=14)

| <i>Variable</i> | <i>Mean (SD)</i> | <i>Kolmogorov-Smirnov test of normality</i> | <i>Skewness</i> | <i>Kurtosis</i> |
|--|------------------|---|-----------------|-----------------|
| Abstraction Score (% eliminated per move) | 34.27 (4.39) | 0.163 | Z=-0.16 | Z=-1.15 |
| Elimination errors (Number of errors) | 6.36 (8.42) | 0.253* | Z=2.47** | Z=6.88** |
| Planning time per move (Seconds) | 33.71 (5.26) | 0.254* | Z=0.352 | Z=-1.402 |
| Delayed Recall (Raw score) | 13.50 (3.78) | 0.226 | Z=0.97 | Z=0.717 |

*Significant deviation from normality at $p < 0.05$ **Significant skew/kurtosis at $p < 0.01$

percent across games for some participants. This indicates that multiple games are necessary as one game taken in isolation could erroneously suggest a deficient or superior performance. Also notable is the highly positive skew of the Elimination Error (EE) variable. This reflects that the majority of participants made very few of these errors (71% made six or less errors for all five games) while a small number made multiple errors, e.g. one participant made 32 elimination errors in total, mainly due to confusing the yes/no response on a number of turns and eliminating the characters that should have been retained.

Figure 3.1 depicts the relationship with age for each variable derived from Guess Who. A Z-score is used for each variable to allow them to be depicted within the same graphs. In the graph at the top of the figure, an expected improvement in GWSS is apparent as participants get older. EE's appear to reduce as expected with increasing age, although this could be due to the aforementioned outlier (age=8.26 years) who made a large number of errors. The patterns in the bottom graph, however is unexpected. Here we see that Delayed Recall score appears to be dropping for older participants.

These patterns were investigated using statistical tests. A significant, moderate positive correlation was observed between Age and GWSS $r(12) = 0.689$, $p=0.003$. A negative relationship between Age and EE's approached significance,

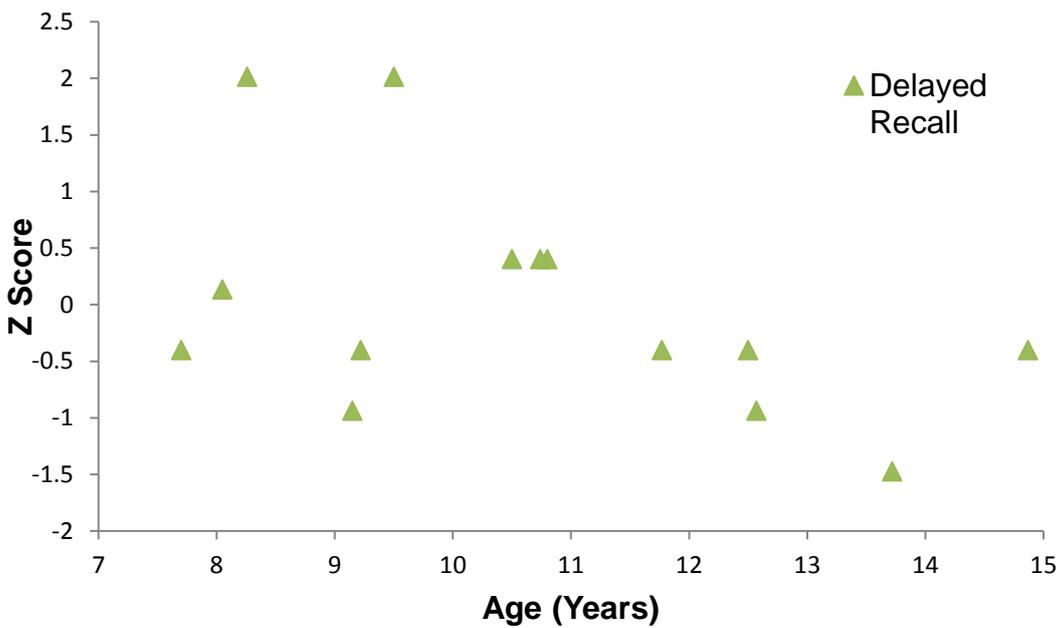
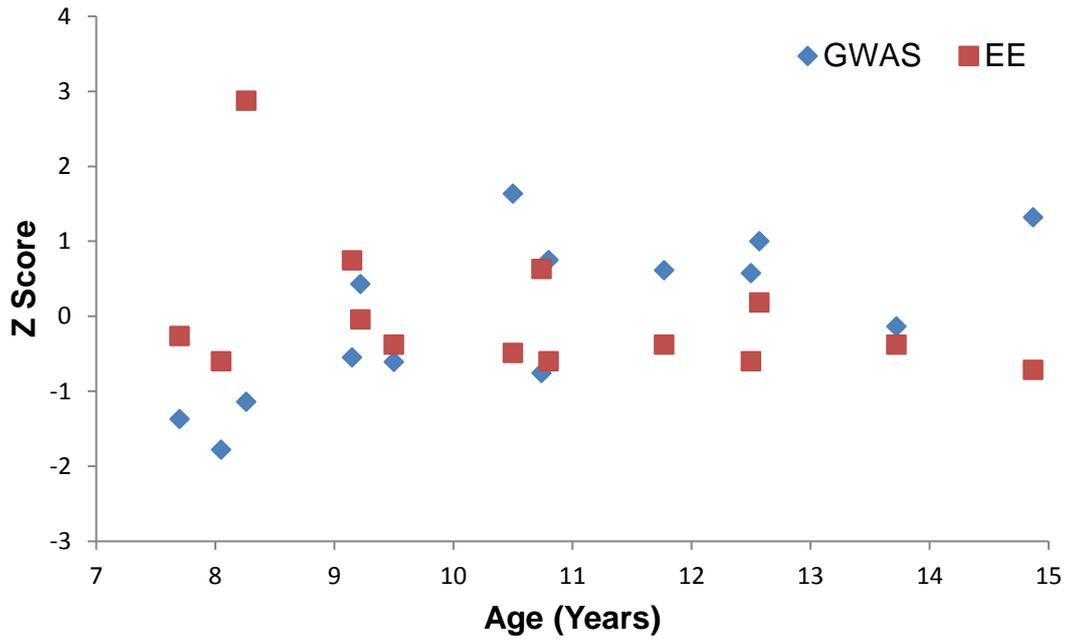


Figure 3.1. Guess Who: Relationship between age and (top) GWSS and EE and (bottom) Delayed Recall.

where the correlation was low $r_s(12) = -0.392$, $p = 0.081$. The observed negative correlation between age and Delayed Recall was low, but reached statistical significance $r(12) = -0.467$, $p = 0.046$, with older participants scoring significantly worse.

To consider the relationship with general intellectual ability, the variables were correlated with WISC-IV Similarities score and RPM score, with an adjusted significance level ($p=0.025$) given the close relationship of both of these measures with full-scale IQ. GWSS showed a high positive correlation with Similarities score $r(12)=0.823$, $p<0.001$ and a moderately high correlation with RPM score $r(11)=0.683$, $p=0.005$. EE's did not show the expected significant negative correlation with Similarities score $r(12)=0.384$, $p=0.088$ or RPM score $r(11)=-0.543$, $p=0.027$ at the adjusted significance level. The correlations were low and moderate respectively. The relationships between Delayed Recall and Similarities score $r(12)=-0.334$, $p=0.121$ and Delayed Recall and RPM score $r(11)=-0.436$, $p=0.068$ were non-significant. However, both correlations were low and negative, which suggests that Delayed Recall may tend to decrease as general intellectual ability increases.

3.5.2 Concurrent validity tests.

The results of the correlations between the variables extracted from Guess Who and the scores on the established neuropsychological tests are shown in Table 3.5. All associations were tested using a Pearson's Product-Moment Correlation test statistic. Statistics relevant to hypotheses are highlighted in the table. The following were the hypotheses and results.

Hypothesis one: GWSS and strategy formation.

The strength of the association between GWSS and 20 Questions Weighted Achievement Score was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two scores. The alternative hypothesis was that a significant positive relationship exists between the two scores.

A statistically significant positive correlation was observed $r(12)=0.643$, $p=0.007$, suggesting that GWSS shows concurrent validity with a valid measure of strategy formation. The correlation coefficient indicated that this was a moderate positive correlation. The coefficient of determination (R^2) indicated that 41.34% of the variance in the scores on the established measure of strategy formation was accounted for by GWSS.

Table 3.5Guess Who: Correlations between scores on tests/games¹ (N=14)

| | <i>TMT A</i> | <i>20 Q's WA</i> | <i>RBANS Del. Recall</i> | <i>Zoo Map Errors²</i> | <i>EE</i> | <i>GWSS</i> |
|-----------------------------------|--------------------|---------------------|--------------------------|-----------------------------------|-------------------|-------------------|
| <i>20-Q's WA</i> | -0.403 (0.076) | | | | | |
| <i>RBANS Del. Recall</i> | -0.549* (0.021) | 0.719** (0.002) | | | | |
| <i>Zoo Map Errors²</i> | 0.351 (0.120) | -0.733** (0.002) | -0.693** (0.004) | | | |
| <i>EE</i> | 0.243 (0.198) | -0.373 (0.094) | -0.612* (0.010) | 0.365 (0.110) | | |
| <i>GWSS</i> | -0.603* (0.011) | 0.643** (0.007) | 0.592* (0.026) | -0.603* (0.015) | -0.397 (0.080) | |
| <i>GW Del. Recall</i> | 0.255 (0.190) | -0.553* (0.020) | -0.417 (0.069) | 0.316 (0.146) | -0.046 (0.438) | -0.284 (0.162) |

¹Pearson's Product-Moment Correlation is shown, except for GW Elimination Errors where Spearman's Rho is shown. Significance level of correlation is in brackets. Shaded figures relate to hypothesised outcomes.

² N=13

*Significant correlation at p<0.05

**Significant correlation at p<0.01

Hypothesis two: GWSS and planning ability.

The strength of the association between GWSS and the number of errors on the Zoo Map Test (Version 1) was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant negative relationship exists between the two scores, i.e. that a better GWSS would be associated with fewer errors on the Zoo Map Test (Version 1). A Bonferroni adjusted p level of 0.025 was used as the GWSS was already used in testing hypothesis one above.

A statistically significant negative correlation was observed $r(12)=-0.603$, $p=0.015$, suggesting that GWSS shows concurrent validity with this measure of planning ability. The correlation coefficient indicated that this was a moderate negative correlation. The coefficient of determination (R^2) indicated that 36.36%

of the variance in the scores on the established measure of planning ability was accounted for by GWSS.

Hypothesis three: EE and visual search and attention.

The strength of the association between EE's and the time to complete the TMT Part A was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant positive relationship exists between the two scores, i.e. that fewer EE's would be associated with a faster completion time on the TMT Part A.

The result did not reach significance $r_s(12)=-0.243$, $p=0.198$. The null hypothesis was therefore accepted. This result suggests that EE's are unlikely to measure visual search and attention skills, at least in how such skills are measured by the TMT Part A. The correlation coefficient indicated that this was a low positive correlation.

Hypothesis four: Guess Who Delayed Recall and recall memory function.

The strength of the association between scores on the Guess Who Delayed Recall and scores on the modified RBANS Figure Recall was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant positive relationship exists between the two scores, i.e. a higher score on Guess Who Delayed Recall would be associated with a higher score on the modified RBANS Figure Delayed Recall.

The result did not reach significance $r(10)=-0.417$, $p=0.069$. The null hypothesis was therefore accepted. This result suggests that Guess Who Delayed Recall is unlikely to function as a valid measure of recall memory function, at least in how such skills are measured by the modified RBANS Figure Recall. Paradoxically, there was a moderate negative relationship between the two variables. This result is perhaps unsurprising given the lack of relationship between Guess Who Delayed Recall and age (see Section 3.5.1).

3.5.3 Fidelity to scoring criteria

Five games of Guess Who were selected randomly from five different participants (one ABI). GWSS and EE were recalculated with the original scoring matrix hidden. The scores awarded to each game were identical. This is unsurprising given the lack of subjectivity in the scoring methodology once the questions have been recorded and eliminations noted.

3.5.4 Practice effects and reliability analysis

All except one of the participants were familiar with the game, having played it several times in the past, although no participant said they had played the game within the previous month. It was notable that the one participant unfamiliar with the game initially struggled and the data from their first game could not be included in the final data set. This raises the issue of familiarity and practice effects and whether participant scores across each of the five games are tapping the same ability. Relatedly, one could query whether the questions asked within each game (1st question, 2nd question, etc.) are also measuring the same underlying construct. This is a potential issue as the number of alternatives reduces on each question, meaning that, for example, visual search skills may be less pressed on later questions.

To investigate this, a reliability analysis was conducted. GWSS across the five games had a relatively low reliability (Field, 2009, p. 675), with Cronbach's $\alpha=0.650$. However, the analysis showed that exclusion of the first game increased the reliability to a high range ($\alpha=0.809$) and that correlations between GWSS on the first game and on subsequent games were either negative or absent. This suggests that GWSS for the first game and subsequent games may measure different underlying constructs, which could relate to practice effects or the relevance of novelty in executive function skills (Gioia & Isquith, 2004), although this would require further investigation. Considering the questions on each game, GWSS across the first four questions had a relatively low reliability, with Cronbach's $\alpha=0.603$. This suggests that each question is not measuring the same underlying construct. This might relate to the fact that GWSS taps executive and visual search and attention skills (See Section 4.3.1 for discussion). The inter-correlation matrix did not show any moderate or high correlations between GWSS on any of the four questions.

3.5.5 Qualitative description of process

Scoring Guess Who in a standardised manner raises some challenges, primarily due to the examiner being unable to see the examinee's characters. The characters that are eliminated by the examinee are recorded by noting with a pen the positions on the board that are turned over on each turn (rather than the names of the eliminated characters). This means that the examiner has no information on the nature of decisions or errors made by the examinee until after the assessment. This in turn reduces the examiner's ability to create hypotheses regarding the examinee's difficulties during the games, which could inform the selection of further tests, or further questions for the examinee or parents regarding the nature of his/her difficulties, etc. This is compounded by the attention required to record the order of elimination and questions asked in an efficient manner in order to not break the flow of the game. This further curbs the examiner's ability to attend to difficulties displayed by the examinee during the game.

3.5.6 Section Summary

The first two hypotheses regarding the variables derived from Guess Who were supported. GWSS showed moderate to high correlations with an established measure of strategy formation (Twenty Questions) and an established measure of planning ability (Zoo Map Version 1 errors). The validity of this variable as a measure of executive function was further supported by the high correlations with age and scores on the two measures of general intellectual ability. However, the narrow standard deviation for the group but wide range of scoring across games for individuals showed that several games are necessary to properly gauge performance. GWSS also showed significant correlations with TMT Part A and with the delayed recall on the modified RBANS. Thus, GWSS correlated with performance on tests of visual search and information processing speed and delayed recall memory. This highlights a lack of discriminant validity for GWSS and, accordingly, how well it selectively measures executive function. Modifications to GWSS that might address this issue are discussed in Section 4.3.1 below.

Hypothesis three was not supported, as EE did not significantly correlate with time to complete an established measure of visual search (TMT A). Similarly, there was no association between Delayed Recall and score on the established measure of delayed recall memory (modified RBANS Figure Recall). Neither EE nor Delayed Recall showed the expected associations with age and general intellectual ability. Higher score on Delayed Recall showed a trend towards an inverse relationship with age and general intellectual ability, which indicates that there were serious flaws with this measure of memory. These are discussed further in the Section 4.3.1 below.

3.6 Connect 4: construct validity

3.6.1 Descriptive statistics and relationship to age and intellectual level

Table 3.6 provides descriptive statistics for each of the variables derived from Connect 4 that were used in the analysis. Each of the three variables displayed properties of normality. Skewness and kurtosis were also within acceptable limits. The number of winning moves and switches available across the five games ranged from between five and 17. If we consider these figures alongside the standard deviations, it seems that these variables provide an opportunity to differentiate between varying levels of performance within five games. For example, consider a case where only five switches each were available to two participants across five games. In this case, if one participant made one less switch than the other it would still place him/her approximately a standard deviation below the other player.

Table 3.6
Connect 4: Descriptive statistics for TD group (N=14)

| <i>Variable</i> | <i>Mean (SD)</i> | <i>Kolmogorov-Smirnov test of normality</i> | <i>Skewness</i> | <i>Kurtosis</i> |
|---------------------------------------|------------------|---|-----------------|-----------------|
| Wins Identified (% identified) | 69.28 (16.90) | 0.154 | Z=-0.106 | Z=-0.819 |
| Switching score (% of switches taken) | 64.84 (22.98) | 0.109 | Z=-0.517 | Z=-0.791 |

*Significant deviation from normality at $p < 0.05$

**Significant skew/kurtosis at $p < 0.01$

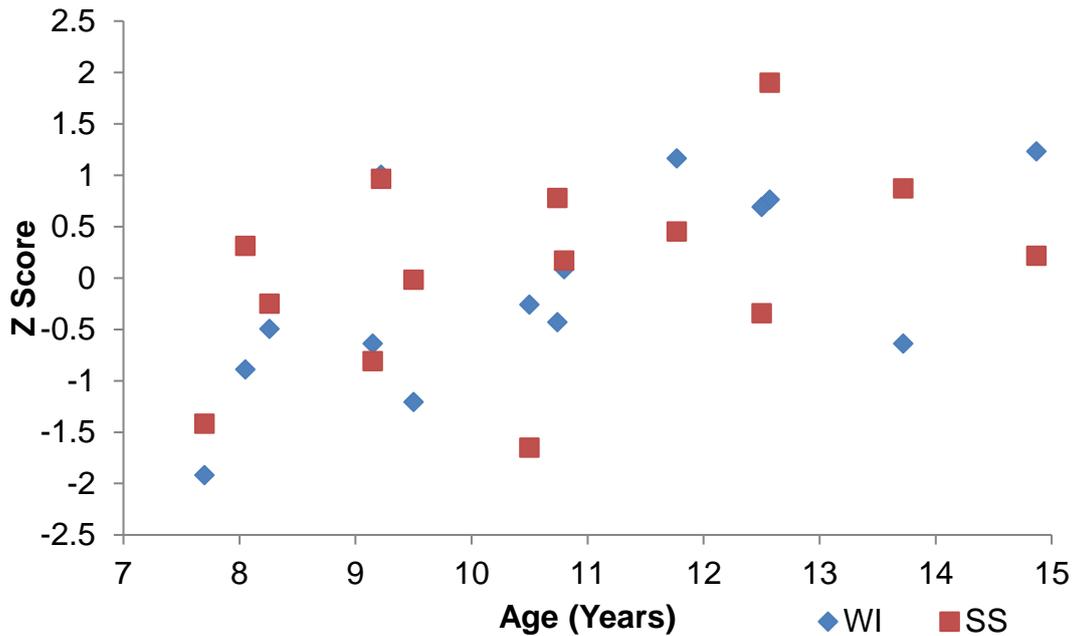


Figure 3.2. Connect 4: graph showing relationship between age and WI / SS.

Figure 3.2 depicts the relationship with age for each variable. Z scores are again used for each variable to allow both to be depicted within the same graph. The graph figure appears to show the expected improvements in Wins Identified (WI) and Switching Score (SS) as participants get older. Statistical tests confirmed this, showing a moderate positive correlation between age and WI that reached significance $r(12)=0.635$, $p=0.007$. The relationship between age and SS was non-significant $r(12)=0.408$, $p=0.074$, although the correlation coefficient indicated that a moderate positive correlation existed between the two variables.

Considering the relationship with the measures of intellectual ability, WI had a significant positive association with both WISC-IV Vocabulary score $r(12)=0.590$, $p=0.013$ and RPM score $r(11)=0.631$, $p=0.010$ and at the corrected significance level. The rule of thumb indicated that both correlations were moderate. No significant relationship was found between SS and Raven's Matrices Score $r(11)=0.108$, $p=0.363$ and WISC-IV Similarities Score $r(12)=0.076$, $p=0.389$.

3.6.2 Concurrent Validity Tests

The results of the correlations between the variables extracted from Connect 4 and the scores on the established neuropsychological tests are shown in Table

Table 3.7Connect 4: Correlations between scores on tests/games¹ (N=14)

| | <i>TMT A</i> | <i>Wins Identified</i> | <i>Switching Score</i> | <i>TMT B-A</i> | <i>MCST Errors</i> |
|--------------------|---------------------|------------------------|------------------------|------------------|--------------------|
| <i>WI</i> | -0.687** (0.003) | | | | |
| <i>SS</i> | -0.370 (0.096) | 0.481* (0.041) | | | |
| <i>TMT B-A</i> | 0.428 (0.064) | -0.502* (0.034) | -0.109 (0.355) | | |
| <i>MCST Errors</i> | 0.238 (0.206) | -0.128 (0.332) | -0.205 (0.240) | 0.055 (0.426) | |

¹Pearson's Product-Moment Correlation is shown, significance level of correlation is in brackets. Shaded figures relate to hypothesised outcomes.

² N=13

*Significant correlation at $p < 0.05$

**Significant correlation at $p < 0.01$

3.7. Statistics relevant to hypotheses are highlighted in the table. The following were the hypotheses and results.

Hypothesis one: WI and visual search and attention.

The strength of the association between WI and time to complete TMT Part A was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant negative relationship exists between the two scores, i.e. that a higher percentage of WI would be associated with a faster completion time for TMT A.

A statistically significant negative correlation was observed $r(12) = -0.687$, $p = 0.003$, indicating that WI displays concurrent validity with this measure of visual search and attention. The correlation coefficient indicated that this was a moderate negative correlation (bordering on a high correlation). The coefficient of determination (R^2) indicated that 47.20% of the variance in the scores on the established measure of planning ability was accounted for by WI.

Hypothesis two: SS and set-shifting.

The strength of the association between SS and errors on the MCST was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant negative relationship exists between the two scores, i.e. that a higher SS would be associated with fewer errors on the MCST.

The correlation failed to reach significance $r(12) = -0.219$, $p = 0.236$, meaning we must accept the null hypothesis. This suggests that SS is unlikely to function as a valid measure of set-shifting, at least in how such skills are measured by the MCST. The correlation coefficient indicated that this was a low negative correlation. The coefficient of determination (R^2) indicated that 4.8% of the variance in errors on the MCST was accounted for by SS.

Hypothesis three: SS and divided attention.

The strength of the association between SS and the adjusted TMT B - A was measured to test this hypothesis. The null hypothesis was that no relationship exists between the two measures. The alternative hypothesis was that a significant negative relationship exists between the two scores, i.e. that a higher SS would be associated with a lower TMT B - A.

The correlation failed to reach significance $r(12) = -0.109$, $p = 0.355$, meaning we must accept the null hypothesis. This suggests that SS is unlikely to function as a valid measure of divided attention, at least in how such skills are measured by the TMT B - A. The correlation coefficient indicated that there was little or no correlation between the variables. The coefficient of determination (R^2) indicated that 1.19% of the variance in TMT B - A was accounted for by SS.

3.6.3 Fidelity to scoring criteria and practice effects

Five games of Connect 4 were selected randomly from five different participants (one ABI). WI and SS were recalculated using the original criteria from the heuristics with the original awarded scores hidden. The scores awarded for WI were identical. For SS, the scores for each game changed, for some participants to a large degree, i.e. for two of the five games the change in the score was greater than the original standard deviation. This indicates that, notwithstanding

the lack of support for Hypotheses two and three, the reliability of SS as a measure is likely to be low.

All participants expressed some familiarity with Connect 4. It appeared that younger participants had more recent experience of the game as it was frequently played at after-schools clubs. Two participants expressed some knowledge of the strategy of playing towards the centre of the board. Their WI scores (83.33% and 88.89%) were towards the upper end of the range of scores. This provides some evidence that experience of the game might affect WI score.

3.6.4 Qualitative description of process

The process of “administering” Connect 4 was markedly easier than for Guess Who. It was relatively undemanding to record each move chronologically while simultaneously monitoring the decisions made by the participant. This could facilitate process observations if used in a clinical setting. For example, missed turns, a lack of attention to the examiner’s repetitive strategy and irrelevant moves could provide further information as to the examinee’s ability to engage with a task requiring close concentration. Although these issues are difficult to codify, they are relatively easy to note while continuing to record performance as was done here.

3.6.5 Section Summary

The first hypothesis, namely that WI is a valid measure of visual search, was supported, with a high negative correlation with TMT Part A observed. WI also correlated highly with age and with the two measures of general intellectual ability, supporting its validity as a neuropsychological measure. The data displayed normality and the spread of scores indicated that WI could be useful in differentiating between varying levels of performance. Considering discriminant validity, the only other measure that WI showed an association with was TMT B - A, which also taps visual search ability. No associations were noted with the established measures of executive function, supporting the conclusion that that WI selectively measures visual search and attention skills.

Hypotheses two and three were not supported. SS did not show an association in the expected direction with the measure of set-shifting or the measure of divided

attention. Also, SS did not show an association with age or general intellectual ability, as would be expected for a measure of executive function. On checking the fidelity to scoring criteria, it was noted that it was not always easy to decide what on what constituted a legitimate switch. This could have contributed to the results found for SS.

3.7 Performance of ABI group

3.7.1 Overview of Group and performance

Given the small number of participants and the confounding factor of age, a group comparison with TD group was unfortunately not possible. Instead, the performance of the ABI participants relative to the means and age-related trends of performance within the TD group are depicted. Where possible, comparison with population normative data was also made. A brief case study of each participant is presented. I will consider the ABI participants' performance on the two variables that have demonstrated promise as valid neuropsychological measures within the TD group, GWSS on Guess Who and WI on Connect 4, and the two established measures they correlate with, D-KEFS 20 Questions Weighted Achievement Score and TMT A time.

The participants with ABI were recruited from a group of patients who did not require an inpatient admission for rehabilitation. Three of the sample had suffered a TBI, one had infant encephalitis and one had suffered a haemorrhage (note that to protect anonymity the pathology cannot be included under the case studies). As outlined in the methods section, this recruitment strategy ensured that all participants were not seriously impaired and could complete the study tasks. All of the ABI participants were in mainstream education without classroom support at the time of testing, some were completing state examinations such as the eleven plus or GCSE's. Thus, differentiating the performance of this group (rather than a severely impaired group) from that of a normative sample is likely to prove relatively challenging.

Table 3.8 provides a profile of ABI participants' performances. Considering age scaled scores where normative data was available, it appears that general verbal ability and performance on the D-KEFS Twenty Questions were well within normal limits. As described below, ABI 1, ABI 2 and ABI 5 performed poorly on the RPM. Unfortunately, normative data for the TMT Part A was not of a suitable age range (Strauss, Sherman & Spreen, 2006, p.762). However, looking at Figure 3.3, it appears that only ABI 3's performance lies outside the range of the TD group.

Figure 3.4 depicts the performance of the five participants on the measures derived from the games alongside the TD group. If one looks at the improvement seen across age within the TD group, there appears to be a trend towards poorer performance within the ABI group. In particular, the same three participants (ABI 1, ABI 2, ABI 5) perform poorly on both measures, with ABI 3 also performing poorly on WI. ABI 4 performance on GWSS and WI is clearly similar to that of TD group.

Interestingly, as noted below, ABI 1, ABI 2 and ABI 5 also scored poorly on the RMT. Also, if we considering the day-to-day difficulties the ABI participants were experiencing as reported by their parents (see Figure 3.5), ABI 1 and ABI 5 also appear to have the most marked difficulties with behaviour regulation. All ABI participants appear to have problems with meta-cognitive control, although ABI 5's problems are once again the most marked.

Table 3.8
Scoring profile of ABI participants

| | Participant ID | | | | | TD mean (SD) |
|---|-----------------|---------------|-----------------|-----------------|---------------|------------------|
| | ABI 1 Female | ABI 2 Male | ABI 3 Female | ABI 4 Female | ABI 5 Male | |
| Age in Years | 12.87 | 14.38 | 14.20 | 12.13 | 11.23 | 10.63 (2.05) |
| Parent's age completing education | 19.50 | - | 18.00 | 23.00 | 16.00 | 20.05 (2.85) |
| WISC-IV Similarities Standard Score | 10 | 11 | 14 | 12 | 11 | - |
| RPM Score/9 | 3 | 2 | 5 | 5 | 2 | 5.18 (2.23) |
| Guess Who GWSS | 28.85 | 33.08 | 39.47 | 35.58 | 29.21 | 35.15 (3.94) |
| Connect 4 WI | 60.00 | 60.00 | 72.73 | 92.86 | 57.14 | 71.20 (17.56) |
| D-KEFS 20 Questions Weighted Achievement score | 3.50 | 3.75 | 2.75 | 4.25 | 3.50 | 3.52 (0.87) |
| D-KEFS 20 Questions Weighted Achievement age SS | 10 | 11 | 7 | 13 | 11 | - |
| TMT Part A | 40.00 | 32.00 | 56.00 | 31.25 | 39.75 | 36.73 (11.83) |

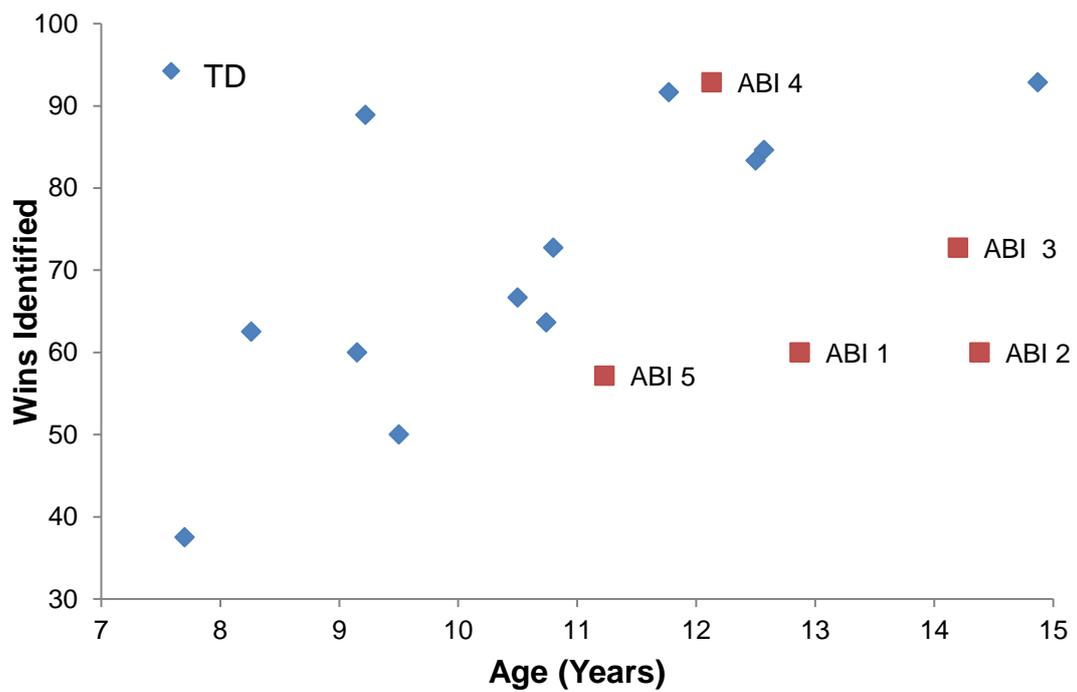
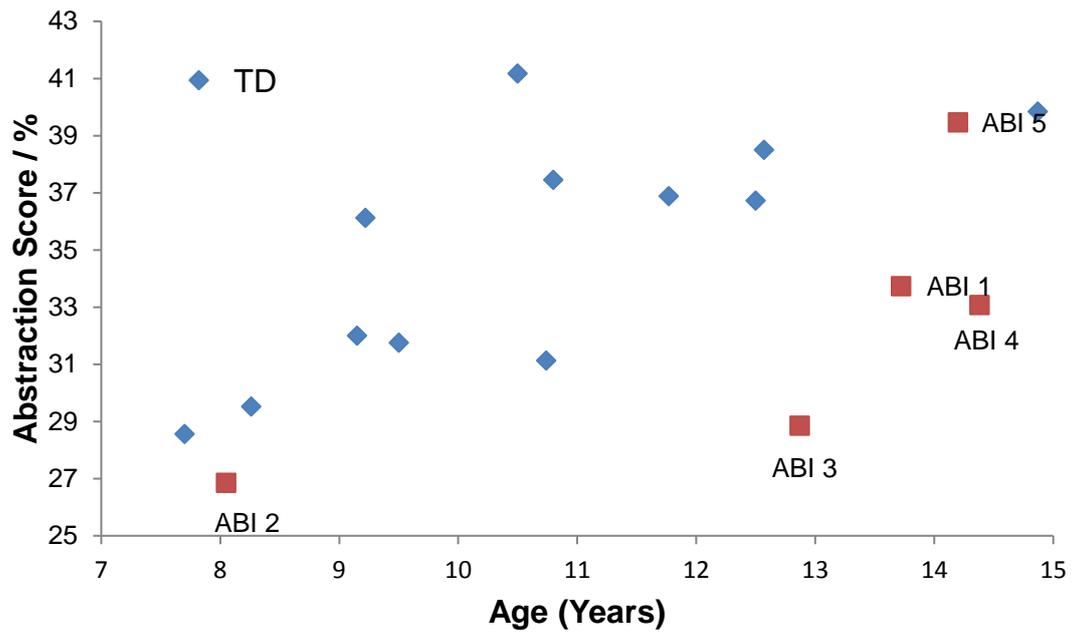


Figure 3.4. Graphs showing relationship between age and (top) GWSS and (bottom) WI with ABI participants highlighted.

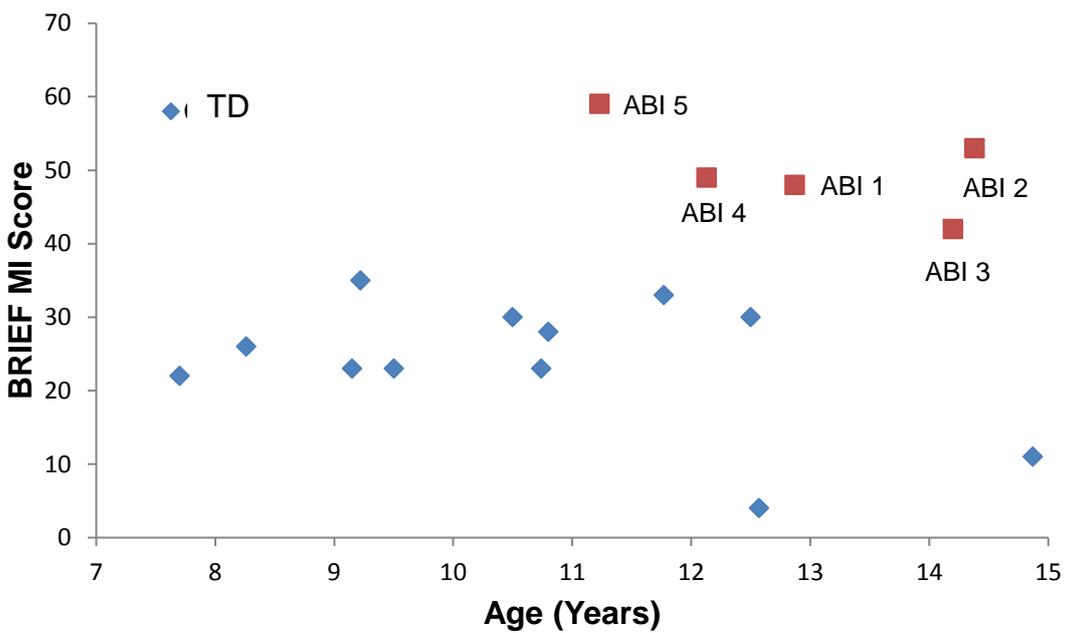
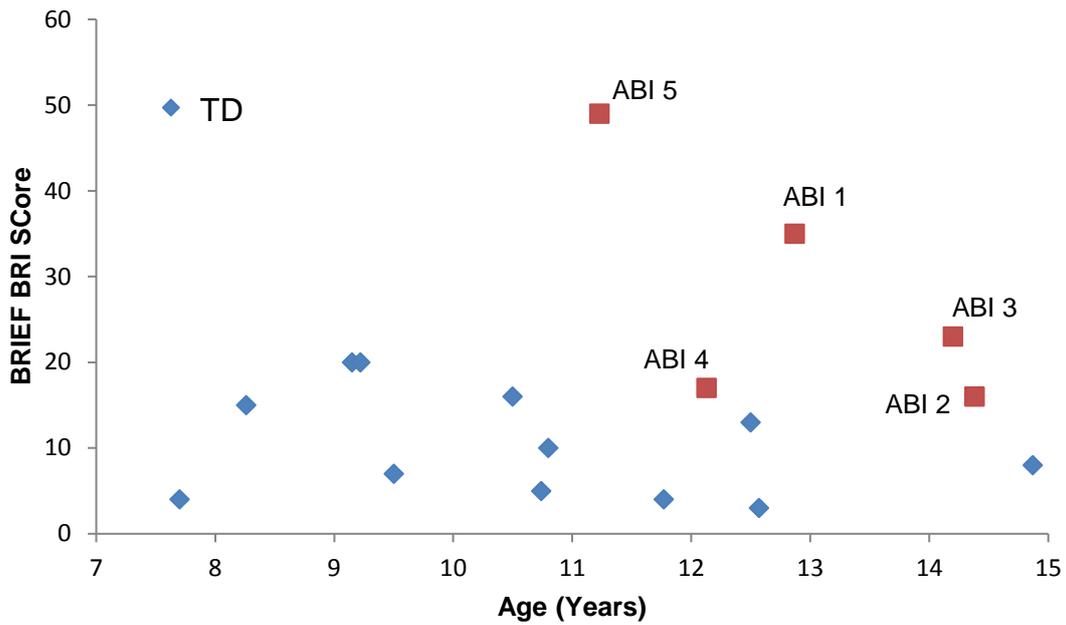


Figure 3.5. Graphs showing relationship between age and (top) BRIEF BRI Score and (bottom) BRIEF MI Score with ABI participants highlighted.

3.7.2 Profile of individual ABI participants' performances

3.7.2.1: ABI 1

ABI 1's parental report on the BRIEF indicated marked difficulties with behavioural regulation and meta-cognitive control. Subjectively, her behaviour

during the games appeared relatively disinhibited and over-familiar and there were some subtle language specific problems evident in her speech. Several repetitions of the examiner's choices of question were noted on Guess Who, which could suggest a range of issues relating to executive function, i.e. with disinhibition, working memory or strategy formation.

Considering the established measures, the most notable result from her profile is a poor performance on the RPM, which would place her in the low average to borderline impaired range, despite a robust score on the measure of general verbal ability (WISC-IV Similarities). Using normative data, her 20 Questions Weighted Achievement Score, by contrast, was in the average range. Her time to complete the TMT A appeared in line with the performance of the TD group (see Figure 3.3).

Her scores on each of the two measures from the games were over one standard deviation below the TD mean, a group who, on average, were over two years younger than her (See Figure 3.4).

3.7.2.2: ABI 2

ABI 2's parental report on the BRIEF indicated concerns with meta-cognition (memory, attention and planning problems) but less concerns with behavioural regulation. Subjectively, his behaviour during the games appeared normal beside some small occasional evidence of boredom. On Guess Who, he repeated questions made by the examiner on four occasions, which again could suggest a range of issues relating to executive function, i.e. with disinhibition, working memory or strategy formation.

Considering the established measures, ABI 2's score on the RPM would place him in the low average to borderline-impaired range relative to the population as a whole. His general verbal ability was estimated in the average range (WISC-IV Similarities) His time to complete the TMT A appeared in line with the performance of the TD group (see Figure 3.3). Using normative data, his 20 Questions Weighted Achievement age standard score was in the average range

ABI 2 reported playing Connect 4 a lot in hospital while recovering from his injury. His WI and GWSS scores were lower than the TD group mean, but within a

standard deviation. However, he was almost four years older than the TD group mean. Figure 3.4 suggests that ABI 2's scores on GWSS and WI were quite weak when the trend of improvement for age is accounted for.

3.7.2.3: ABI 3

ABI 3's parental report on the BRIEF indicated concerns with meta-cognitive control and behavioural regulation. Subjectively, no major difficulties were noted by the examiner with ABI 3's behaviour or thinking skills while playing the games, although some frustration at performance was evident.

Considering the established measures, ABI 3's scores appeared, overall, robust. General verbal ability was above average (WISC-IV Similarities Score) and RPM score suggests that non-verbal ability is also in the average range. Her extremely slow time on the TMT Part A relative to the TD group (see Figure 3.3) is likely an anomaly as her time on the objectively more difficult TMT Part B was faster. 20 Questions Weighted Achievement Score was a little weak but still broadly within the average range using normative data for comparison.

ABI 3's GWSS scores were higher than the TD group average and markedly better than her WI Score (see Figure 3.4). The relative involvement of language skills in Guess Who and ABI 3's strength in this domain may have underpinned this result.

3.7.2.4: ABI 4

ABI 4's parental report on the BRIEF indicated concerns with meta-cognitive control but relatively less concern for behavioural regulation. ABI 4 did not exhibit any overt difficulties with her behaviour or cognition while playing the games.

ABI 4's scores on all of the established measures were robust. Her general verbal and non-verbal ability (WISC-IV Similarities, RPM) were in the average range as was her 20 Questions Weighted Achievement Score. Her time to complete TMT Part A was slightly quicker than the normative group, although she was approximately 1.5 years older than the TD group average (see Figure 3.3).

Both ABI 4's GWSS and WI scores appeared comparable to that of the TD group of a similar age (see Figure 3.4). Her WI score was over one standard deviation above the TD group average.

3.7.2.5: ABI 5

ABI 5's parental report on the BRIEF indicated the greatest concern for difficulties with both meta-cognitive control and behavioural regulation out of all of the participants. Subjectively, he was somewhat distractible during testing and asked questions of the examiner that could be seen as relatively socially inappropriate.

ABI 5's verbal intelligence appeared relatively robust (WISC-IV Similarities: average range) relative to a weak non-verbal performance ability (RPM: low average to borderline-impaired range). His 20 Questions Weighted Achievement score was also in the average range. Examination of the TD group mean and inspection of Figure 3.3 appears to indicate that his TMT A time is broadly within the expected range.

ABI 5's GWSS and his WI Score were below the group means, almost one standard deviation below for GWSS and over one standard deviation below for WI (see Table 3.8). Thus, his apparent strength in the verbal domain did not translate into a GWSS score comparable to the TD group.

3.7.3 Section Summary

Although a group comparison was not viable given the low group numbers, and participants with ABI were considerably older than the TD group, the group data and case summaries provide some indication that GWSS and WI might differentiate ABI participants from TD CYP. Particularly, in comparison with the 20 Questions measure and TMT A time, GWSS and WI appeared to place ABI participants below the mean performance of the TD group. If the trend of improvement with age for the TD group on GWSS and WI is considered, this weakness in performance for ABI's appears even more apparent. Also, the participants with the greatest difficulties as per a parental report and those with apparent difficulties in non-verbal ability tended to score the worst on the GWSS and WI. Overall, a larger normative sample to allow robust comparisons at each age range with TD CYP is required. This could allow conclusions to be drawn

regarding the ability of GWSS and WI to discriminate CYP with ABI from age-matched TD CYP.

3.8 Ecological Validity

3.8.1 Comparison of experimenter's observations and parental ratings

Figure 3.6 depicts the ratings of parents for each participant on the proxy report with the experimenter's ratings of observed difficulties embedded within the graph. Unsurprisingly, greater difficulties with metacognition and behavioural regulation were reported for the ABI group. Due to the low number of participants and the lack of age-matching between the groups, a statistical comparison of the ABI and TD groups was not possible. However, the psychometric properties of the proxy report, which has narrow percentile ranges across stratified age groups (Gioia et al., 2000), allow an investigation of the usefulness of the experimenter's observations.

As the experimenters ratings of strengths and difficulties were influenced by whether the participant had an ABI or not, both groups are represented separately. Observations of strengths/difficulties were based on breakdown in performance of the game or enhancement of performance through the recruitment of appropriate strategies. Observations of general behaviour that impinged upon gameplay also informed observations. For the ABI group, difficulties in the area of behavioural regulation were observed for two participants. The difficulties noted were problems with turn-taking during the games, socially inappropriate comments towards the experimenter (personal compliments), getting up abruptly to get food, asking the experimenter to play other games and a general over-familiarity that was seen as age-inappropriate. One participant was rated as having a strength in metacognition due to her ability to sustain attention in the presence of reported fatigue and potential distracters. For the TD group, strengths in behavioural regulation were noted for four participants who were attuned to the context of the testing procedure and the need to eliminate distracters. In terms of metacognition, one participant that recalled target characters from previous games and used this to eliminate alternative was rated as having a strength in this area. Weaknesses in working memory were noted for some participants in the form of repeated questions

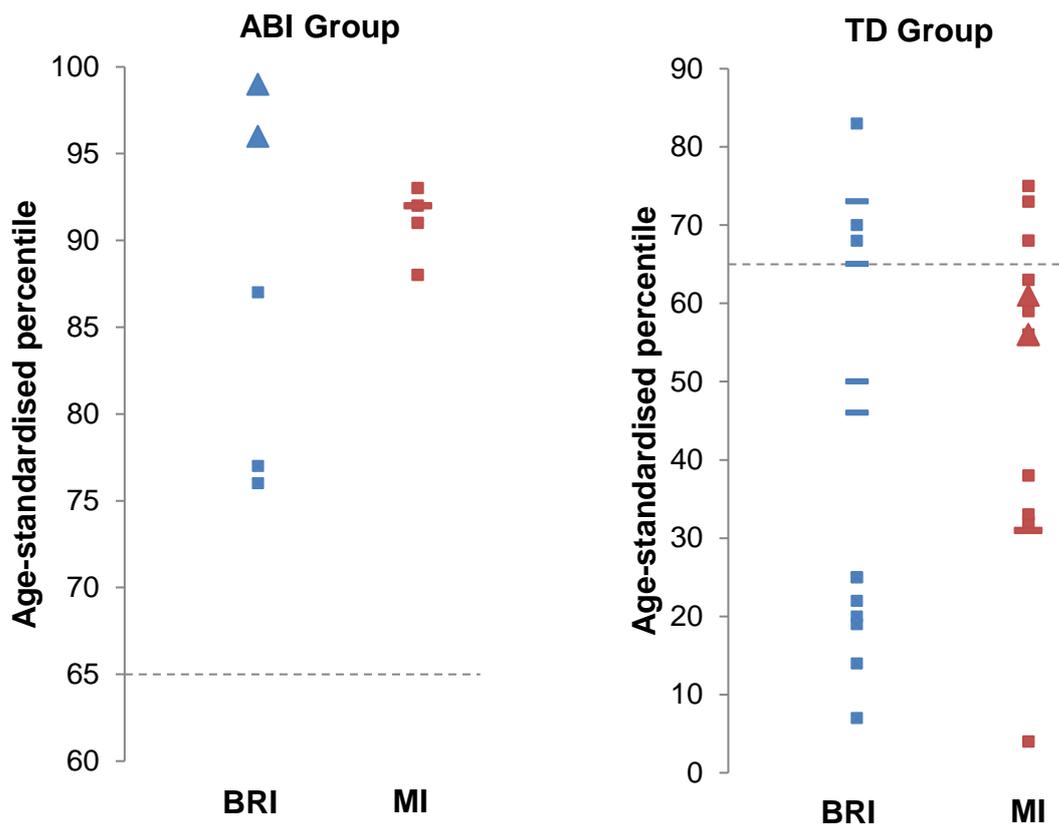


Figure 3.6: Graphs depicting parent-rated difficulties for individual participants on the BRIEF for the ABI group (left) and TD group (right). A higher percentile represents greater reported difficulties. Participants represented with a dash were rated by the experimenter as having strengths on the index in question. Conversely, participants represented with a triangle were rated by the experimenter as having a weakness on the index in question. Ratings came from observations while participants played the games. BRI = Behavioural Regulation Index, MI = Metacognition Index. Dashed line indicates cut-off for clinically significant difficulty (above the line) as per author’s manual. Note that each participant is represented once in the BRI column and once in the MI column.

As can be seen from Figure 3.6, the ratings based on the observations of the examiner were relatively unsuccessful in categorising difficulty as per the parental reports. For the ABI group, the two participants rated by the examiner as having behavioural regulation problems were indeed the two participants with the highest parental ratings in this area. However, the participant rated by the examiner as having a strength in the area of metacognition was not distinguishable from their

peers with ABI in this area, with parental reported difficulties well into the area of clinical significance. For the TD group, those rated with strengths in the area of behavioural regulation by the examiner were indistinguishable from the broader TD group on the parental report. One of these participants' ratings falling within the clinical range. For metacognition, neither of the TD participants rated as having potential difficulties by the examiner fell within the clinical range as per parental report and their ratings were in the middle of the range for the TD group as a whole. The TD participant rated by the experimenter as having strengths in metacognition did demonstrate the second lowest parental rating for difficulties within this area. However, the percentile score of the parental rating (31st percentile) is relatively unremarkable relative to their peer group from the population as a whole.

3.8.2 Ecological validity of GWSS and WI

As discussed in the introduction, there is evidence for a lack of ecological validity for traditional, quantitative neuropsychological measures. Therefore, it is worth investigating whether the two quantitative measures derived from the games that show promise in terms of validity will show an association with the parental reports of difficulties. For GWSS, no association was found either with the BRIEF behavioural regulation index $r(12)=0.177$, $p=0.272$, or the BRIEF metacognition index $r(12)=0.107$, $p=0.358$. Similarly, for WI, no association was found either with the BRIEF behavioural regulation index $r(12)=0.168$, $p=0.283$, or the BRIEF metacognition index $r(12)=0.068$, $p=0.405$.

3.8.3 Section summary

This section indicates that the ecological validity of Guess Who and Connect 4 as a brief neuropsychological screen appears weak. Although the numbers in TD and ABI groups were small, observations of difficulties while playing Guess Who and Connect 4 did not seem to bear a relationship with executive function difficulties as found on a standardised parental report. Also, no association was found between GWSS and WI and ratings on the parental report. Some potential reasons for these results are discussed in full in Section 4.3.5.

4. Discussion

4.1. Section Overview

The following discussion focuses on how the results relate to the strengths and weaknesses of the study design and methods. The design of the measures derived from Guess Who and Connect 4 are firstly discussed, focussing on how issues of construct validity could be addressed. The findings regarding engagement and ecological validity are then considered, particularly in light of the profile of the sample and the shortcomings of the quality of observations offered by the method. A broader discussion highlights conceptual issues with the gamification of tests with regards to neuropsychological theory and clinical practice.

4.2 Summary of main results

For Guess Who, GWSS (a measure of the quality of game-playing strategy) showed concurrent validity with established tests of executive function. GWSS also correlated highly with age and general intellectual ability, supporting the construct validity of this measure given the relationship between executive function and general ability (Duncan et al., 2000). However, correlations between GWSS performance on a test of visual search and attention and a test of delayed recall memory suggests that the discriminant validity of GWSS may be weak. Scores fell within a narrow range, meaning that a relatively minor change in score could result in a large change in estimated ability. Omission of scores from the first game improved the reliability of GWSS, possibly suggesting that the first game is affected by additional factors such as practice effects. Two further measures were derived from Guess Who, EE (the number of characters mistakenly kept or eliminated) and Delayed Recall (ability to recall position of characters on the board) did not show concurrent validity with established measures.

For Connect 4, WI (the number of winning moves taken or blocked) showed concurrent validity with a test of visual search and attention. WI correlated highly with age and general intellectual ability, lending support to the construct validity of the measure. There was evidence of discriminant validity for WI, as no

association was found between WI and performance on established tests of executive function. The range of WI scores was reasonably wide, indicating that it could potentially differentiate between a range of levels of performance.

Switching Score (SS, the ability to change between building and blocking a line) showed no association with performance on tests of executive function and divided attention or with age and general intellectual ability. A difficulty obtaining the same SS when rescored games of Connect 4 indicated issues with the reliability of this measure.

Considering other results, a statistical analysis showed that participants with ABI and TD participants did not rate the games as more enjoyable than the tests. Completing both the games and the tests evoked little anxiety across the entire sample. Three participants with ABI performed particularly poorly on GWSS and WI relative to TD participants. They also had weak scores on a measure of non-verbal ability and the worst functional executive function difficulties as per the parental reports. Considering ecological validity, the experimenter's ratings of participants' executive function skills based on observations during the games did not reflect those reported by parents on a proxy report.

4.3 Implications of results

4.3.1 Design of measures: Guess Who

The results indicated that GWSS has potential as a screening measure for CYP with ABI. GWSS appeared to measure executive function and visual search and attention skills. Tests that load highly on these constructs, such as the TMT, have been found to be sensitive to mild and moderate ABI in general in CYP (Reitan & Wolfson, 2004) and adults (Chan et al., 2015). Unfortunately, the structure of the games would make it difficult to tell from GWSS whether either visual search or executive function is more affected by an ABI. This could limit its usefulness as a screening tool as a clinician might struggle to create hypotheses regarding the exact nature of a client's cognitive deficits.

It was also noted that GWSS correlated with delayed recall of a complex figure, which also limits its discriminant validity. Two possible explanations for this can be considered. Firstly, GWSS might tap such a range of cognitive functions that it

also correlates with memory ability. A second and perhaps more cogent explanation is based on the observation that encoding the complex figure relies on organisational ability and visual construction skills. In adults, it is known that frontal lobe lesions disrupt the ability to encode the figure correctly by disrupting these organisational and construction skills (Messerli, Seron & Tissot, 1979). Thus, it might be the case that the encoding stage of this memory test was also tapping an element of executive function, which could explain the correlation with GWSS. However, as the figure copying score was not entered into the correlation matrix this explanation remains a conjecture.

Test design could be modified to increase the ability of GWSS to discriminate executive function difficulties. One relevant issue with Guess Who was that the ability to eliminate characters after each question removed the need to keep past questions “in mind”. This reduces the load on working memory. This is significant given the very strong relationship between working memory and all executive function skills (McCabe et al., 2010). If working memory demands were increased, one would assume that the proportion of variance in GWSS due to executive function skills would increase and, potentially, make GWSS load relatively less on more elementary visual search skills. Forcing participants to view all of the faces throughout the game by dropping the ability to eliminate characters after questions might achieve this.

Addressing issues with stimulus design might also enhance the ability of Guess Who to differentiate executive function skills from other cognitive skills. It is notable that the attributes participants needed to identify in faces were relatively concrete, e.g. gender, hair colour, etc. Contrast this with the D-KEFS Twenty Questions Test, where the shared attributes include “being alive” or “being machinery”. These more abstract verbal concepts increase the need for cognitive flexibility associated with executive function (Ardila, Pineda & Roselli, 2000). Therefore, making the shared attributes in Guess Who more abstract might also increase how much executive function is engaged. Characters could have similar concrete physical characteristics but wear clothing which indicates that they work in different fields (i.e. healthcare, business) or display facial expressions that are more subtle (i.e. pensive, suspicious). A mix of more concrete shared attributes

could be retained to ascertain whether an advantageous strategy would be followed whether abstract verbal reasoning is more or less taxed.

Game design might also explain the inability of EE to measure visual search skills. Within established tests of visual search, the influence of variables other than those affecting the accurate identification of the target is held constant. For example, in the TMT Part A, the targets are digits. As digits have a simple form, difficulties perceiving more complex visual forms do not affect score. As reciting the number line seems conceptually undemanding, the effect of language skill is also reduced. With Guess Who, however, such wider variables impinge on the visual search task. Any issues with face perception, for example, would completely compromise EE as a measure of general visual search skills. Some examinees may also be less familiar with subtle descriptions of a character's hair colours (e.g., brown, ginger, blonde) for cultural reasons. A judgement as to whether a character looked happy or merely content could also be quite subjective or reliant on language skill. Therefore, EE could have been influenced by linguistic ability or cultural factors. Creating a more culture-fair set of stimuli might address some of these criticisms. Facial expressions could be selected that satisfy the criteria of Matsumoto (1992) for portraying universal emotions. The ethnicity of the characters could also be broadened. The current set featured four characters of African background, with the rest entirely White European. A character set more representative of the wider population might reduce the effect of shared physical attributes that are more common in people of White European background. However, even if such changes are made, retaining EE would mean retaining the examinee's ability to eliminate characters during the game. As just noted this affects working memory load within the task and potentially the ability of GWSS to discriminate executive function difficulties. Therefore, any future modification to the methods used here might require a choice between retaining either GWSS or EE.

Guess Who Delayed Recall score showed particularly poor validity, with an apparently negative relationship with age and general intellectual ability. These results could be related to a few issues with the structure of the test. Firstly, participants were not explicitly reminded to familiarise themselves with the faces, their names or their locations. This is common in visual memory tests to ensure

that a more general factor such as attention does not underpin performance (e.g. Warrington, 1996). It may have been that younger and less able participants scored better as a result of this; their attention to the central task may have been poorer, allowing them to pick up on less task-relevant information such as face location. A second issue was that no gestalt exists between the faces, which were placed randomly on the board. Such an underlying structure is apparent in stimuli for visual recall tasks and facilitates encoding of the stimulus for later retrieval (Shin et al., 2006). Thirdly, even if the above two issues were addressed, the central task (i.e. GWSS) involves a focus on a verbal description of the characters, which could be at odds with engaging visuo-spatial skills to encode character location. Clearly, addressing the marked shortcomings of the delayed recall measure would require large modifications to Guess Who game-play, which in turn is likely to interfere with processes being measured with GWSS and EE.

4.3.2 Design of measures: Connect 4

WI's correlation with the measures of visual search and attention (but not the tests of executive function) shows its potential as a screening test for ABI. As already discussed, the trails tests are sensitive to brain injury in general; in adults sensitivity to mild injury following TBI has been demonstrated (Lange et al., 2005). Connect 4 was easy to administer compared to Guess Who and, if SS is not assessed, obtaining the WI score would be a quick and easy task. Also, given the abstract nature of Connect 4, one can speculate that WI would be a relatively culturally-fair test. In fact, given that attention to numerals in visual space is differentially affected by whether the number line is represented left to right (e.g. English) or right to left (e.g. Arabic) in one's native language (the "snarc effect", Zebian, 2005), it is likely that WI would be a more culturally-fair test than both TMT A and TMT B.

Basic problems with the SS construct may have underpinned the difficulties confirming the score when revisiting previously scored games. Referring to Figure 2.6, one can ascertain that assessing whether a player is genuinely switching requires some inference. One example noted was the following. One could allow the opponent to continue building an apparently winning line. While

the opponent is occupied with this strategy, one could assemble a winning move that is unassailable once the opponent is later blocked. Assessing whether a player was indeed pursuing such a strategy did require some inference regarding the participant's intentions. Another issue related to whether game-play did in fact require one to hold in mind the need to switch from building to blocking. Participants may have learned that placing tokens as close as possible to the centre of the rack increases one's options of building a winning line and blocking the experimenter. If this strategy was adopted successfully by many of the participants, either from previous experience or from experience gained during the study, the need to switch or "shift set" may have been minimal.

An issue noted while playing both games was how the estimation of a participant's ability affected the strategy adopted by the experimenter. In Guess Who, ensuring that the first three questions were invariant across participants was straightforward. It was also relatively easy to adopt a more or less advantageous strategy after these three questions to attempt to finish the game at or around the same time as the participant. However, a difficulty was encountered if the experimenter was "lucky" with a poor question. For example, if the experimenter eliminated 20 alternatives with his initial question, it would make sense for the participant to adopt a riskier strategy in order to win. Despite this strategy being sensible, it would result in a lower GWSS. For Connect 4, this issue was more complicated still. Having set moves for the first two or three tokens was again straightforward. After this, the decisions made to prolong the game and facilitate assessment were more complex. For example, often a win would be made available or intentionally missed to see whether these would be identified by the participant. However, doing this required some guess-work as to the participant's ability; if a participant was quite able then making easy wins available could result in a ceiling effect and a difficulty estimating ability.

This issue highlights a general difference between neuropsychological tests and two-player games. It is relatively easy to increase the difficulty in standard neuropsychological tests, for example, by making targets in a visual search task more distinctive (e.g. compare the Bells Test, Gauthier et al, 1989 with the Star Cancellation Test, Halligan, Wilson & Cockburn, 1991). Encoding can also be facilitated in memory tests by making stimuli more easy to encode (e.g. compare

Pictorial and Topographical Memory Tests, Warrington, 1996). However, varying how advantageous an opponent's strategy is within a game appears to be a relatively more complex task, one that is likely to require collaboration with disciplines outside of psychology such as computer science (see varying difficulty levels for Connect 4 on Maths Is Fun, 2014). The wider implications of this issue for the gamification of neuropsychological tests are discussed in Section 4.4 below.

A more prosaic issue raised by the involvement of the examiner in dynamically setting task difficulty is that the examiner's strategies may tend to confirm their subjective impressions of the examinee's cognitive difficulties. This is particularly relevant to the current study as the examiner was not blind to the presence of ABI.

4.3.3 Engagement with games and tests

The analysis showed that participants did not find completing the games significantly more enjoyable than completing the tests and experienced little anxiety while completing both the games and the tests. While this might appear to question the usefulness the games have in increasing engagement with testing, some caveats should be noted. Firstly, the clinical observations that led to this study were made in acute or inpatient settings. Mood problems are common in CYP with ABI (Schwartz et al., 2003) and one could presume that acute and inpatient settings are by their nature more likely to increase such difficulties due to, e.g. multiple demands for medical and rehabilitative sessions, disorientation and concerns regarding prognosis. Perhaps the value of the games in increasing engagement may only become apparent in the presence of such heightened anxiety and lowered mood. Secondly, it is worth considering the nature of the sample in the study. All participants volunteered for the study having been given a description of what participation would involve. They were presumably unperturbed by the prospect of completing cognitive testing or else would have declined to participate. The sample is therefore probably unrepresentative of what would be encountered in an acute or inpatient setting. Testing in such settings is probably more likely to be seen as "mandatory" for the sake of rehabilitation regardless of mood or anxiety. Perhaps such a sample

might benefit in assessment from the increased levels of engagement that games have already demonstrated in rehabilitation work (Perry et al., 2011).

4.3.4 Ability of games to discriminate ABI and TD group performance

The data appeared to indicate that the performance of many ABI participants lagged that of TD participants on WI and GWSS, which indicates some promise for these measures as screening tests. The observation that the ABI and TD participants appeared to perform similarly on other established measures, e.g. the tests of general intellectual ability, could support the hypothesis that WI and GWSS are particularly sensitive to the presence of ABI.

Clearly any such conclusions would require further study using a between-subjects design with matched ABI and TD groups. Such a study could show whether GWSS and WI reliably differentiate the performance of CYP with cognitive deficits due to ABI from their TD peers. It could also allow cut-offs to be derived and validated to help understand the sensitivity and specificity of WI and GWSS in categorising the performance of CYP with ABI. Knowledge of this is key to establishing the clinical value of a screening measure in identifying cases requiring intervention and those where further testing is likely to be burdensome for the patient and service (Wilson & Jungner, 1968).

4.3.5 Ecological validity of testing method

From the data in the study, it appears that performance on Guess Who and Connect 4 are not reliable indicators of reported real-world executive function difficulties. This was true using either in-session observations of game-play or scores on the quantitative measures derived from the games. There are a number of potential reasons for this result. These include the difficulty making observations while “administering” the games, the structure provided to participants by the games, the absence of concrete criteria against which to compare observations and the impact of parental anxiety on rated difficulties. These are now discussed along with their clinical implications.

Gioia and Isquith (2004) note the clash between traditional validity considerations and concerns for ecological validity. In the current study, interaction with the experimenter in the form of turn-taking and general interpersonal interaction may

have provided insight into the nature of executive function difficulties. However, as deriving reliable measures from the games took close concentration for 45 minutes, opportunities for the examiner to observe the effects of such intervening variables was limited. A future attempt to assess cognition using play could embed a briefer measure of cognition within a longer period of play. For example, embedding a brief, structured visual construction task within a wider observation of play with bricks, lego, etc. could allow for more open-ended interaction and observation. A related recommendation is that the measures embedded within an ecologically valid task should be easily recordable to facilitate observation of behaviour during testing.

Intact performances on tests of executive function are frequently present alongside real-world difficulties (e.g. Stuss & Buckle, 1992). An example of this was participant ABI 4, who performed well on the GWSS measure, but their difficulty with real-world tasks as per the proxy report fell well into the clinical range. To understand this, one can consider the constraints that were placed on the CYP when playing Guess Who. This included the structure of the rack that allowed alternatives to be easily eliminated and the explicit instruction to only use yes/no questions. This may have facilitated the focussed, quantitative assessment of executive functions. However, as Silver (2000) points out, such structure reduces the need for self-monitoring and coping with novelty, two key issues for CYP with real-world executive function difficulties. Thus, the structures that allowed performance to be codified for the quantitative measure may in fact have undermined the verisimilitude of the test.

Another relevant issue was raised by Silver (2000), who describes the inherent difficulty in creating reliable scoring criteria for ecologically valid measures. For example, she notes that lengthy observations are often required for proxy reports to ensure all items can be reliably responded to. In the current study, it was (in retrospect) unlikely that data from observations would augment the quantitative measures derived from the games given the short duration of testing. This issue with setting outcome criteria is addressed in observed real-world tasks by creating specific scenarios to (potentially) force performance to break down. For example, in the Children's Cooking Task (Chevignard et al., 2010) being distracted by deliberately inserted variables, failing to monitor performance based

on a final goal or omitting a step in a series of instructions are coded as errors. The absence of such specific process tasks to relate difficulties to when playing Guess Who or Connect 4 may have made it difficult to identify and categorise behaviours relevant to executive function problems.

It was notable that participants with ABI were more likely to be categorised on the parental proxy report as having severe difficulties with executive function. This was despite all of them attending mainstream schools without extra support and scoring at or above average on the WISC-IV Similarities subtest. The latter is a demanding test of abstract verbal reasoning that correlates significantly with executive function skills (Ardila, Pineda & Roselli, 2000). Although this test may simply not pick up the difficulties noted on the proxy report, it is also possible that the presence of an ABI might increase parental anxiety and thus vigilance for difficulties (Chevignard et al., 2012). Alternatively, parents may have become accustomed to emphasising difficulties on such questionnaires as part of acquiring necessary support, e.g. in school, for their children. The negativity scores on the proxy report for the ABI participants lend some support to these ideas. This has two potential implications for a clinician completing ecologically valid tests. Firstly, a strong score on a traditional measure alongside parental reports of severe problems should raise questions of ecological validity, but also the potential issue of diagnostic overshadowing (Reiss, Levitan, & Szyszko, 1982). This could mean discussing the findings of proxy reports with parents and understanding if observed difficulties reflect cognitive effects of ABI, mood or adjustment difficulties for the CYP, or perhaps even difficulties that were present premorbidly but have become more apparent due to parental anxiety and vigilance. A second issue is that any ecologically valid assessment should incorporate tasks with a verisimilitude with tasks that are a concern for the CYP and their parents. It was noted that issues completing school assignments and studying for examinations were a major noted concern for parents and participants in the current study. Guess Who and Connect 4, however, may have had greater verisimilitude with difficulties with social interaction, e.g. turn-taking, cooperation, which could have been of less concern to parents.

A general point arising from the analysis of ecological validity concerns the use of in-session observations by clinicians to help guide formulations of cognitive

difficulties. The use of process observations has been described as a key part of information gathering in the neuropsychological assessment of CYP (Warner-Rogers & Reed, 2008, p. 436). These observations have been explicitly recommended as an avenue of understanding how executive function difficulties (from test results) manifest for CYP in real-world settings (Hughes & Graham, 2008, p.268). However, the number of studies supporting these recommendations seems small. There is some evidence that test session observations are a useful aid in assessing and identifying ADHD (Glutting, Robins & De Lancey, 1997; McConaughy et al., 2009). However, these studies examined the association between task-general difficulties observed in sessions (e.g., inattention, opposition) and outcomes on cognitive assessments. Neither study demonstrated that these observations could predict real-world difficulties related to specific cognitive domains. The latter was also not possible in the current study, which suggests that the usefulness of in-session observations might be limited. However, it must be acknowledged that in usual clinical practice session observations are likely to be coupled with wider observational evidence, previous reports and opinion from other multi-disciplinary team members. Despite this, the data here highlights how an individual clinician's interpretive bias might enter into the testing room, something possibly more openly acknowledged in other areas of clinical psychology practice.

4.3.6. Viability of method as a neuropsychological screen

Using Guess Who and Connect 4 in the manner used here does have potential to screen CYP for deficits following confirmed or suspected ABI. The two issues highlighted by Guess Who and Connect 4, namely attention and executive function, are highly susceptible to TBI, the most common form of ABI in CYP. These are accordingly seen as a key area for assessment in CYP (Ylvisaker & Gioia, 1998). Relative to the methods reviewed in the introduction, Guess Who and Connect 4 appear to provide greater observational data than the simple pen-and-paper speed and attention tasks that were used for screening, even though the value of these observations was not confirmed here. Issues were shown with the discriminant validity of the games, particularly with GWSS. This indicates that the method explored here is less likely to identify specific cognitive deficits compared with screening tools containing brief tasks covering multiple domains.

In terms of screening, Guess Who and Connect 4 also required several iterations to produce a reliable measure with a reasonable range of scores. This issue would need to be addressed to demonstrate their ability to function as a “brief” screen.

4.3.7 Further study limitations

Some further limitations are worth noting that could affect the ability to generalise results.

Firstly, the sample size was clearly small, and although the post-hoc analysis showed that it was suitably powered, the group were almost entirely recruited from White, English speaking, middle-income families in the South East of England. Therefore the income and ethnic profile of the participants is not representative of the UK as a whole. The familiarity of, for example, minority ethnic groups with Guess Who and Connect 4 and the likely effect of this on testing with them cannot therefore be commented on.

Secondly, another important effect of the small sample size was the low number of participants at specific age ranges. An increased sample within narrow age ranges would be vital in creating valid cut-off scores for impairment across the childhood and adolescence. This could also help identify specific areas of performance that improve across the age range. It is known that the executive function component of working memory develops relatively early, which could underpin the later-emerging ability to hold strategies in mind and modify them (see Best & Miller, 2010). If elimination of characters was dropped as described in Section 4.3.2 , with a large sample size one could provide data on both components via the assessment of repeated questions and strategy errors.

A third issue concerns practice effects. It was notable that the only participant expressing no familiarity with Guess Who made a number of errors on the first game, which made the data from this game uninterpretable. However, their performance improved markedly on subsequent games. This highlights how familiarity with the games may have influenced performance. A structured questionnaire of how often and recently participants had played both games could have helped understand any such familiarity. A further note is that the

reliability analysis indicated that exclusion of the first game of Guess Who from analysis greatly increased reliability of GWSS. This suggests that the first game of Guess Who could be used to simply familiarise examinees with the format of the game, which could also help eliminate the effects of familiarity.

A fourth issue concerns the study design. If one is to infer that a causal role in behaviour can be attributed to an internal cognitive/neural structure, such a structure needs to demonstrate a causal role within the study. GWSS and WI correlated with the established measures, but also with age and general intellectual ability. It might have been that these tasks were just difficult and therefore more easily completed by older, more able participants. One way to demonstrate a causal role would be via an experimental design (Shadish, Cook & Campbell, 2002, pp. 1-32). For example, participants with known, specific difficulties with executive function could be compared with participants with e.g. known, specific memory problems on GWSS. If such groups were matched as closely as possible for age and general intellectual ability, the outcome could indicate whether intact neural structures relating to executive function do in fact have a causal role in outcome on GWSS.

A final issue relates to the measures chosen to estimate general intellectual ability. As already outlined, WISC-IV Similarities and RPM both correlate highly with full-scale IQ. However, similar to GWSS, the structure of both tests calls for the identification of higher-level concepts, verbal in the case of Similarities (Lezak et al., 2012, p. 621) and non-verbal/visual in the case of RPM (Lezak et al., 2012, p. 629). The RPM is a noted measure of fluid intelligence (Raven, Raven & Court, 2003), which is known to correlate highly with all executive function measures (Roca et al., 2010). It could be argued that these two measures were too similar to GWSS to allow an estimation of the true relationship between GWSS (or any other executive function measure) and general intellectual ability. Using these two measures to measure general intellectual ability also might provide a misleading picture for participants with ABI who had specific deficits in concept formation and fluid intelligence. Administration of a full test of intellectual ability that included subtests of more crystallised ability, such as the WISC-IV (Wechsler, 2004), could have remedied this.

4.4 General discussion

Potential difficulties generalising the findings could be understood based on the epistemological position adopted. The critical realist position assumed ontological realism, i.e. that the neural structure of participant's brains effected performance. However, the interpretive lens through which behaviour was categorised, i.e. executive function, attention, is viewed more as socially constructed. This has clear implications for a correlational study design where, as just discussed, the purported cause cannot be directly manipulated to determine its effect on another variable (Shadish et al., 2002, pp. 1-32). The idea that this executive function is somewhat socially constructed has some support. Executive function development is affected by cultural factors such as an emphasis on self-control (Lewis et al., 2009) and bilingualism (Carlson & Meltzoff, 2009). Emotional and social development are also known to underpin executive function development (Diamond & Lee, 2011), which fits with the known effects of deprivation and trauma on frontal lobe development (see Gerhardt, 2015, for extensive review). Therefore, particularly for GWSS, any conclusions regarding the generalisability of the findings requires replication of the results within a sample raised in a different cultural milieu.

This issue also calls forth an element of reflexivity in understanding the methods used here to measure cognitive abilities (Yardley, 2000). The chief investigator was a white, European male. The idea that a game would naturally call forth competitive behaviour could be seen as a judgement synonymous with a Western achievement-oriented culture, and perhaps even a gendered judgement. It is interesting to reflect that social cognition development, closely interlinked with executive function development (Blakemore & Choudhury, 2006), is clearly required to facilitate cooperative behaviour. This was not considered when designing the study, which might have reflected the researcher's bias that healthy executive function might entail "out-witting" the researcher in a game of strategy rather than engaging cooperatively with the researcher. Cultural expectations of what "controlled behaviour" entails should thus be considered in future studies of executive function that incorporate an element of observation.

The disconnect observed between test scores and real-world functioning raises questions regarding clinicians' roles and future test development. The previous importance of neuropsychological testing in localising lesions and aiding diagnostic investigations has diminished with technological advances in medical imaging. However, this does not automatically mean that a clinician's role should involve describing functional deficits, something already possible with occupational therapy measures (e.g., AMPS, © Center for Innovative OT Solutions). The identification of higher level constructs informed by brain function that help categorise behaviours into syndromes remains central to the value of a neuropsychologist's work. It is not clear whether increasingly opting to observe and categorise real-world difficulties would lead to advances in the development of such constructs. One potential route of linking the measurement of higher-level constructs with predictions of functional deficits may lie by increased consideration of whole-brain or connectivity models. For example, the HERA model could help inform how an executive function deficit sits within the complex process of memory encoding and retrieval (Habib, Nyberg & Tulving, 2003). An understanding of the Default Mode Network could help understand how an executive function deficit could result in the brain being unable to suppress specific sensory information from other brain areas (Uddin et al., 2009). A potential advantage of such models versus the lesion-deficit approach is that they could provide a clinician with more specific hypotheses as to how disrupted "executive function" might affect wider cognitive, emotional and sensory process. This in turn could help build a richer understanding of what an examinee's real-world difficulties look like and how they are underpinned by whole-brain function.

The results of this study also raise a question regarding the practicality and relevance, if any, of the gamification of tests for neuropsychological testing. As already discussed, games generally involve a dynamic strategy on the part of both players. Thus, using tests as games takes neuropsychological testing into the area of dynamic cognitive assessment (see Grigorenko & Sternberg, 1998 for overview). Such an approach, which focusses on what emerges in the interaction between the examiner and the examinee, is potentially at odds with the project of understanding internal brain function.

Technology could address this issue by eliminating the need for the examiner to create a strategy during the game. As already discussed, online algorithms can vary the difficulty of games to a fine level of complexity. This is not easily achieved by an examiner, who usually only has the choice of easier or more difficult tests and an understanding of the examinee's ability to ensure that the data interrogates the true level of ability. Guess Who or Connect 4 presented on a tablet computer could progressively increase or decrease the quality of the examinee's opponent's strategy. Using a staircase method (see Blake & Sekuler, 2005, pp. 558-559) one could arrive at a reliable score for the level at which the examinee can perform. The clinician could still play the game, albeit while being provided their next move by the algorithm. A further advantage of this method is that it would greatly reduce the burden of data collection on the examiner. This would allow the examiner to focus more attention on observing the process of the assessment rather than recording the examinees responses. A downside to this approach, however, is the added expense of the testing method, something that an efficient screening method would seek to avoid.

4.5 Conclusion

This study provided evidence for the concurrent validity of two measures derived from Guess Who and Connect 4. Wins Identified (WI) on Connect 4 showed good promise as a measure of visual search and attention. The measure was easy to administer and relatively culture-fair. Guess Who Strategy Score (GWSS) also showed promise as a measure of executive function. However, further refinement of the measure is required to enhance discriminant validity. Administration of GWSS and WI is likely to tap difficulties commonly associated with ABI in CYP. This indicates that, taken together, both measures could form an initial screening tool for CYP with suspected ABI. This was given some mild support by the case studies of five CYP with ABI, where poor performance on WI and GWSS relative to the TD group was common. Further investigation of the value of the gamification of tests is required with a larger sample. However the data here do not seem to support the use of this testing method to provide insight into the nature of functional difficulties. This method also requires the examiner to respond to changes in examinee game-play, which creates difficulties in deriving

reliable quantitative measures of cognitive function and thus in detecting underlying behavioural syndromes.

5. References

- Airaksinen, E., Larsson, M. & Forsell, Y. (2005). Neuropsychological functions in anxiety disorders in population-based samples: evidence of episodic memory dysfunction. *Journal of Psychiatric Research*, 39, 207–214.
- Anderson, V., Lajoie, G., & Bell, R. (1997). *Neuropsychology assessment of the school-aged child*. Department of Psychology, Royal Children's Hospital, Melbourne, Australia.
- Anderson V, & Moore C. (1995). Age at injury as a predictor of outcome following pediatric head injury. *Child Neuropsychology*, 1,187-202.
- Anderson V., Spencer-Smith, M. & Wood A. (2011). Do children really recover better? Neurobehavioural plasticity after early brain insult. *Brain*, 134(8), 2197–2221.
- Ansari, D. & Karmiloff-Smith, A. (2002). Atypical trajectories of number development: a neuroconstructivist perspective. *Trends in Cognitive Sciences*, 6(12), 511–516.
- Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology*, 15, 31–36.
- Baddeley, A., Emslie, H. & Nimmo-Smith, I. (1994) *Doors and People*. Bury St Edmunds, Thames Valley Test Company.
- Baldo, J. V., Delis, D., Wilkins, D., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Archives of Clinical Neuropsychology*, 19, 407–419.
- Bates, E., Reilly, J., Wulfeck, B., Dronkers, N., Opie, M. et al. (2001). Differential effects of unilateral lesions on language production in children and adults. *Brain and Language*, 79(2), 223-265.
- Bechara, A. (2007). *Iowa gambling task professional manual*. Lutz: Psychological Assessment Resources

- Berg, C., Edwards, D. F., & King, A. (2012). Executive function performance on the children's kitchen task assessment with children with sickle cell disease and matched controls. *Child Neuropsychology*, *18*(5), 432-448.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641-1660.
- Bhaskar, R.A., (1997). *A Realist Theory of Science*. London: Verso.
- Bilker, W.B., Hansen, J.A., Brensinger, C.M., Richard, J., Gur, R.E. & Gur, R.C. (2012) Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*, 354–369.
- Billard, C., Vol, S., Livet, M.O., Motte, J., Vallee, L., & Gillet, P. (2002₁). The BREV neuropsychological test: Part I: Results from 500 normally developing children. *Developmental Medicine and Child Neurology*, *44*, 391–398.
- Billard, C., Motte, J., Farmer, M., Livet, M.O., Vallée, L., Gillet, P. et al. (2002₂) The BREV neuropsychological test. Part II. Results of validation in children with epilepsy. *Developmental Medicine and Child Neurology*, *44*, 398-404.
- Blake, R. & Sekuler, R (2005). *Perception*. London: McGraw-Hill Higher Education
- Blakemore, S.J. & Choudhury, S. (2006) Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology & Psychiatry*, *47*, 296–312.
- Brooks, B. L., Sherman, E. M. S., Iverson, G., Slick, D. J., & Strauss, E. (2010). Psychometric foundations for the interpretation of neuropsychological test results. In M. R. Schoenberg & J. G. Scott (Eds.), *The Black Book of Neuropsychology: A Syndrome-Based Approach*. London: Springer.
- Carlson, S. M., & Meltzoff, A. N. (2008). Paper: Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282-298.

- Carr, A. (2006) *The handbook of child and adolescent clinical psychology. 2nd Edition*. London: Routledge, 2006.
- Casey, R., Ludwig, S. & McCormick, M.C. (1986). Morbidity following minor head trauma in children. *Pediatrics*, 78, 497–502.
- Cattelani, R., Lombardi, F., Brianti, R. & Mazzucchi, A. (1998). Traumatic brain injury in childhood: intellectual, behavioural and social outcome into adulthood. *Brain Injury*, 4, 283-296.
- Chan, E., MacPherson, S. E., Robinson, G. et al. (2015). Limitations of the trail making test part-B in assessing frontal executive dysfunction. *Journal of the International Neuropsychological Society*, 21(2), 169–174.
- Charvet, L.E., Beekman, R., Amadiume, N., Belman, L.A., & Krupp, L.B. (2014). The Symbol Digit Modalities Test is an effective cognitive screen in pediatric onset multiple sclerosis (MS). *Journal of the Neurological Sciences*, 341, 79–84.
- Chevignard, M. P., Catroppa, C., Galvin, J., & Anderson, V. (2010). Development and evaluation of an ecological task to assess executive functioning post childhood TBI: The children's cooking task. *Brain Impairment*, 11(2), 125-143.
- Chevignard, M. P., Soo, C., Galvin, J., Catroppa, C., & Eren, S. (2012). Ecological assessment of cognitive functions in children with acquired brain injury: A systematic review. *Brain Injury*, 26(9), 1033-1057.
- Chomsky, N. (1959). Review of Skinner's Verbal Behavior. *Language*, 35, 26-58.
- Cianchetti, C., Corona, S., Foscoliano, M., Contu, D., & Sannio-Fancello, G. (2007). Modified Wisconsin Card Sorting Test (MCST, MWCST): Normative Data in Children 4-13 Years Old, According to Classical and New Types of Scoring. *Clinical Neuropsychology*, 21(3), 456-478.
- Cohen, D. (2006). *The Development of Play*. UK, Hove: Routledge.

- Coughlan, A.K., Oddy, M. & Crawford, J.R. (2007). *Birt memory and information processing battery*. London: Brain injury rehabilitation trust, Wakefield.
- 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*, 2037–2078.
- Delis, D.C., Kaplan, E. & Kramer, J.H. (2001). *The Delis-Kaplan Executive Function System*. San Antonio, TX: Psychological Corporation.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old (science (959)). *Science*, *334*(6054), 311.
- Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., et al. (2000). A neural basis for general intelligence. *Science*, *289*, 457-460.
- Engel-Yeger, B., Josman, N., Rosenblum, S. (2009). Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C): An examination of construct validity. *Neuropsychological Rehabilitation*, *19*, 662–676.
- Emslie, H., Wilson, C., Burden, V., Nimmo-Smith, I., Wilson, B. (2003). *Behavioural Assesment of Dysexecutive Syndrome for Children (BADS-C)*. UK, Titchfield: Thames Valley Tests Company.
- Epsy, K.A., Kaufmann, P.M., McDiarmid, M.D., et al (1999). Executive functioning in preschool children: performance on A-not-B and other delayed response format tasks. *Brain and Cognition*, *41*, 178– 199.
- Eslinger, P., Biddle, K., & Grattan, L. (1997). Cognitive and social development in children with prefrontal cortex lesions. In N. Krasnegor, G. Lyon, & P. Goldman-Rakic (Eds.), *Development of the prefrontal cortex: Evolution, neurobiology, and behavior* (pp. 295–336). U.S., Baltimore: Brookes Publishing.
- Ewing-Cobbs, L., Prasad, M. R., Kramer, L., Cox, C. S., Baumgartner, J., Fletcher, S., et al. (2006). Late intellectual and academic outcomes following

- traumatic brain injury sustained during early childhood. *Journal of Neurosurgery*, 105(4 Suppl), 287–296.
- Ewing-Cobbs, L., Kramer, L., Prasad, M., et al. (1998). Neuroimaging, physical, and developmental findings after inflicted and noninflicted traumatic brain injury in young children. *Pediatrics*, 102, 300-307.
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.
- Field, A. (2009). *Discovering Statistics Using SPSS*. London: Sage.
- Flanagan, D.P. & Kaufman, A.S. (2004). *Essentials of WISC-IV Assessment*. UK, Chichester: Wiley.
- Folstein, M. F, Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental State': A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 12, 189-198.
- Friston, K.J. (2011). Functional and effective connectivity: a review, *Brain Connectivity*, 1(1), 13-36.
- Gauthier, L., Dehaut, F. & Joanett, J. (1989). The Bells test. A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology*, 11, 49–54.
- Gerhardt, S. (2015). *Why Love Matters (2nd Edition)*. London: Routledge.
- Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadeif, H., Rizzo, A.S. & Josman, N (2011). Using a Virtual Classroom environment to describe the attention deficits profile of children with Neurofibromatosis type 1. *Research in Developmental Disabilities*, 32, 2608–2613.
- Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, 25(1-2), 135-158.

- Gioia, G.A., Isquith, P.K., Guy, P.K. & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function (BRIEF)*. US, Odessa, FL: Psychological Assessment Resources.
- Glosser, G., Zvil, A.S., Glosser, D.S., et al. (2000). Psychiatric aspects of temporal lobe epilepsy before and after temporal lobectomy. *Journal of Neurology, Neurosurgery & Psychiatry*, 68, 53–58.
- Glutting, J. J., Robins, P. M., & De Lancey, E. (1997). Discriminant validity of test observations for children with attention deficit/hyperactivity. *Journal of School Psychology*, 35(4), 391-401.
- Grafman, J., Jonas, B., & Salazar, A. (1990). Wisconsin Card Sorting Test performance based on location and size of neuroanatomical lesion in Vietnam veterans with penetrating head injury. *Perceptual & Motor Skills*, 71, 1120-1122.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124(1), 75-111.
- Habib, R., Nyberg, L., & Tulving, E. (2003). Hemispheric asymmetries of memory: The HERA model revisited. *Trends in Cognitive Sciences*, 7(6), 241-245.
- Halligan, P., Wilson, B., & Cockburn, J. (1990). A short screening test for visual neglect in stroke patients. *International Disability Studies*, 12(3), 95-99.
- Hinkle, D.E., Wiersma, W. & Jurs, S.G. (2003). *Applied Statistics for the Behavioral Sciences (5th ed.)*. Boston: Houghton Mifflin.
- Headway (2015). *Acquired Brain Injury: The numbers behind the hidden disability*. Downloaded from <https://www.headway.org.uk/about-brain-injury/further-information/statistics/statistics-resources/>.
- Higginbotham, D. (10th January, 2012). *An Exhaustive Explanation of Minimax, a Staple AI Algorithm*. Retrieved from <http://www.flyingmachinestudios.com/programming/minimax/>

- Hughes, C & Graham, A. (2008). Executive Functions and Development. In J. Reed & J. Warner-Rogers (Eds.) *Child Neuropsychology: Concepts, Theory, and Practice* (pp. 264-284). John Wiley and Sons, NY.
- Irwin, E.C. (2000). The Use of a Puppet Interview to Understand Children. In K. Gitlin-Weiner, A. Sandgrund & C. Schaefer (Eds). *Play diagnosis and assessment, (2nd Edition)*, (pp. 682-703). US, Hoboken, NJ: John Wiley & Sons.
- Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child*, 2, 125-132.
- Kahnemann, D. (2011). *Thinking, Fast and Slow*. London: Penguin.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental approach to cognitive science*. Cambridge, MA: MIT Press.
- Kolb, B. & Whishaw, I.Q. (2003). *Fundamentals of Human Neuropsychology* (5th Edition). New York: Worth Publishers.
- Kortte, K.B., Horner, M.D., & Windham, W.K. (2002). The trail making test, part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, 9(2), 106–109.
- Krull, K.R., Okcu, M.F., Potter, B., Jain, N., Dreyer, Z., Kamdar, K. et al. (2008). Screening for neurocognitive impairment in pediatric cancer long-term survivors. *Journal of Clinical Oncology*, 26, 4138–43
- Lange, R. T., Iverson, G. L., Zakrzewski, M. J., Ethel-King, P. E., & Franzen, M. D. (2005). Interpreting the trail making test following traumatic brain injury: Comparison of traditional time scores and derived indices. *Journal of Clinical and Experimental Neuropsychology*, 27(7), 897-906.
- Langlois, J.A., Rutland-Brown, W., Wald, M.M. (2006). The epidemiology and impact of traumatic brain injury: a brief overview. *Journal of Head Trauma Rehabilitation*, 21, 375–378.

- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology, Human Perception and Performance*, 21, 451–468.
- Lebby, P.C., Pollock, M., Mouanoutoua, A. & Lewey, J.H. (2014). Performance of Children and Adolescents With Brain Injury on the Lebby-Asbell Neurocognitive Screening Examination for Children and Adolescents. *Journal of Child Neurology*, 30(10), 1255-1262.
- Levin, H.S., Culhane, K.A. & Mendelsohn, D. (1993). Cognition in relation to magnetic imaging in head injured children and adolescents. *Archives of Neurology*, 50, 897–905.
- Lewandowski, L.J. (1984). The Symbol Digits Modalities Test: A screening for brain-damaged children. *Perceptual and Motor Skills*, 59(2), 615–618.
- Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B., & Huang, Z. (2009). Culture, executive function, and social understanding. *New Directions for Child and Adolescent Development*, 2009(123), 69-85.
- Lezak, M.D., Howieson, D.B. Bigler, E.D. & Tranel, D. (2012). *Neuropsychological Assessment, Fifth Edition*. Oxford University Press.
- Lin, C.C.H., Hsiao, C.K., & Chen, W.J. (1999). Development of sustained attention assessed using the continuous performance test among children 6–15 years of age. *Journal of Abnormal Child Psychology*, 27, 403–412.
- Longaud-Valès, A., Chevignard, M., Dufour, C., Grill, J., Puget, S., et al. (2016). Assessment of executive functioning in children and young adults treated for frontal lobe tumours using ecologically valid tests. *Neuropsychological Rehabilitation*, 26(4), 558-583.
- Mark Rober (2015, 17th November). *BEST Guess Who Strategy- 96% WIN record using MATH* [video file]. Retrieved from: <https://www.youtube.com/watch?v=FR1bNOno5VA>.
- Maths Is Fun, Four In A Line (2014). Retrieved from <https://www.mathsisfun.com/games/connect4.html>

- Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, 23, 72–84.
- McConaughy, S. H., Ivanova, M. Y., Antshei, K., & Eiraldi, R. B. (2009). Standardized observational assessment of attention deficit hyperactivity disorder combined and predominantly inattentive subtypes. i. test session observations. *School Psychology Review*, 38(1), 45-66.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24, 222–243.
- Messerli, P., Seron, X., & Tissot, R. (1979). Quelques aspects des troubles de la programmation dans le syndrome frontal. *Archives Suisse de Neurologie, Neurochirurgie et de Psychiatrie*, 125, 23-35.
- MIT (2010). *Connect 4, March 9, 2010*. Retrieved from <http://web.mit.edu/sp.268/www/2010/connectFourSlides.pdf>.
- Mioshi E., Dawson K., Mitchell J., Arnold R., & Hodges, J.R. (2006). The Addenbrookes Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21, 1078–1085.
- Music, G. (2016). *Nurturing Natures: Attachment and Children's Emotional, Sociocultural and Brain Development* (2nd Edition). Hove: Psychology Press.
- Nasreddine, Z.S., Phillips, N.A., Bédirian V, Charbonneau, S., Whitehead, V. ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society*, 53, 695– 699.
- Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12, 313-324.

- Ouvrier, R.A., Goldsmith, R.F., Ouvrier, S. & Williams, I.C. (1993). The value of the Mini-Mental State Examination in childhood: A preliminary study. *Journal of Child Neurology*, 8, 145–148.
- Olson, K., Jacobson, K.K., Van Oot, P. (2013). Ecological validity of pediatric neuropsychological measures: current state and future directions. *Applied Neuropsychology: Child*, 2, 17–23.
- Pejnovic, L. P., De Luca, C. R., Gentle, E., Anson, K., Ashley, D. M., Anderson, V. A., & McCarthy, M. C. (2012). Feasibility of neurobehavioral screening following diagnosis of pediatric cancer. *Pediatric Blood & Cancer*, 59(2), 295-300.
- Perry, J.C., Andureu, J., Cavallaro, F.I., Veneman, J.F., Carmien, S.P., and Keller, T. (2010). Effective game use in neurorehabilitation: user-centered perspectives. In P Felicia (Ed.) *Handbook of Research on Improving Learning and Motivation through Educational Games* (pp. 1-44). IGI Global, US: Hershey, PA.
- Pind, J., Gunnarsdottir, E.K., & Johannesson, H.S. (2003). Raven's Standard Progressive Matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, 34, 375-386.
- Power, J.T. & Radcliffe, J. (2000). Assessing the Cognitive Ability of Infants & Toddlers Through Play: The Symbolic Play Test. In K. Gitlin-Weiner, A. Sandgrund & C. Schaefer (Eds). *Play diagnosis and assessment, (2nd Edition)*, (pp. 58-79). US, Hoboken, NJ: John Wiley & Sons.
- Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function*, 1–38. Hove: Psychology Press
- Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status*. Psychological Corporation, San Antonio.
- Raven, J., Raven, J.C., & Court, J.H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

- Reed, J. & Warner-Rogers, J. (2008) Introduction. In J. Reed & J. Warner-Rogers (Eds.) *Child Neuropsychology: Concepts, Theory, and Practice* (pp. 1-5). John Wiley and Sons, NY.
- Reiss, S., Levitan, G. W., & Szyszko, J. (1982). Emotional disturbance and mental retardation: Diagnostic overshadowing. *American Journal of Mental Deficiency, 86*(6), 567-574.
- Reitan, R.M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271-276.
- Reitan, R. M. (1971). Trail Making Test results for normal and brain-damaged children. *Perceptual and Motor Skills, 33*, 575–581.
- Reitan, R.M. & Wolfson, D. (2004). The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology, 19*, 281–288.
- Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie tramatique. *Archives de Psychologie, 28*, 21.
- Rober, M. (17th November 2015). *BEST Guess Who Strategy- 96% WIN record using MATH*. [Video File]. Retrieved from <https://www.youtube.com/watch?v=FR1bNOno5VA>.
- Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., et al. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain, 133*, 234–47.
- Roy, A., Allain, P., Roulin, J-L, Fournet, N., & Le Gall, D. (2015) Ecological approach of executive functions using the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C): Developmental and validity study. *Journal of Clinical and Experimental Neuropsychology, 37*(9), 956-971.
- Sacks, O. (2015). *The Man Who Mistook His Wife for a Hat*. London: Picador.
- Sánchez-Cubillo, I., Periáñez, J.A., Adrover-Roig, D., Rodríguez-Sánchez, J.M., Ríos-Lago, M., Tirapu, J., et al. (2009). Construct validity of the Trail Making

- Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438–450.
- Schwartz, L., Taylor, H. G., Drotar, D., Yeates, K. O., Wade, S. L., & Stancin, T. (2003). Long-term behavior problems following pediatric traumatic brain injury: Prevalence, predictors, and correlates. *Journal of Pediatric Psychology*, 28(4), 251-263.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shin, M.S., Park, S.Y., Park, S.R., Seol, S.H. & Kwon, S.J. (2006). Clinical and empirical applications of the Rey–Osterrieth Complex Figure Test. *Nature Protocols* 1, 892–899.
- Silver, C.H. (2000). Ecological validity of neuropsychological assessment in childhood traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 15, 973–988.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary (3rd ed.)*. New York: Oxford University Press.
- Stuss, D. T., & Buckle, L. (1992). Traumatic brain injury: Neuropsychological deficits and evaluation at different stages of recovery and in different pathologic subtypes. *Journal of Head Trauma Rehabilitation*, 7(2), 40-49.
- Tate, R.L. (1998). “It is not only the kind of injury that matters, but the kind of head”: the contribution of premorbid psychosocial factors to rehabilitation outcomes after severe traumatic brain injury. *Neuropsychological Rehabilitation*, 8, 1–18
- Toussaint-Thorin, M., Marchal, F., Benkhaled, O., Pradat-Diehl, P., Boyer, F. & Chevignard, M. (2013). Executive functions of children with developmental

dyspraxia: Assessment combining neuropsychological and ecological tests. *Annals of Physical and Rehabilitation Medicine*, 56(4), 268-287.

Uddin, L. Q., Kelly, A. M. C., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2009). Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Human Brain Mapping*, 30(2), 625-637.

Upton, D., & Thompson, P.J. (1999). Twenty questions task and frontal lobe dysfunction. *Archives of Clinical Neuropsychology*, 14, 203–216.

Warner-Rogers, J. & Reed, J. (2008) A Clinician's Guide to Child Neuropsychological Assessment and Formulation. In J. Reed & J. Warner-Rogers (Eds.) *Child Neuropsychology: Concepts, Theory, and Practice* (pp. 432-449). John Wiley and Sons, NY.

Warrington, E.K. (1996). *The Camden Memory Tests*. Psychology Press, Brighton, UK.

Wechsler, D. (2004). *The Wechsler intelligence scale for children—fourth edition*. London: Pearson.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, 7, 131–149.

White, D. A., Moinuddin, A., McKinstry, R. C., Noetzel, M., Armstrong, M., & DeBaun, M. (2006). Cognitive screening for silent cerebral infarction in children with sickle cell disease. *Journal of Pediatric Hematology/Oncology*, 28(3), 166-169.

Wilson, B.A., Alderman, N., Burgess, P., & Emslie, H. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St. Edmunds (UK): Thames Valley Test Company.

Wilson, J.M.G. & Jungner, G. (1968) Principles and practice of screening for disease. *Public Health Papers*, 34. Geneva: World Health Organization.

- Yardley, L. (2000). Dilemmas in qualitative health research. *Psychology and Health, 15*(2), 215-228.
- Ylvisaker, M. & Gioia, G.A. (1998) Cognitive assessment. In: Ylvisaker M, (Ed.) *Traumatic Brain Injury Rehabilitation: Children and Adolescents*. 2nd ed. (pp. 159–179) Boston: Butterworth-Heinemann.
- Ylvisaker, M., Jacobs, H.E. & Feeney, T. (2003). Positive supports for people who experience behavioural and cognitive disability after brain injury: a review. *Journal of Head Trauma Rehabilitation, 18*, 7–32
- Yochim, B., Baldo, J. & Nelson, A. & Delis, D.C. (2007). D-KEFS Trail Making Test performance in patients with lateral prefrontal cortex lesions. *Journal of the International Neuropsychological Society, 13*, 704–709.
- Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: the SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates and illiterate Arabic speakers. *Journal of Cognition and Culture, 5*, 165–191.

Appendix A: Systematic literature review of neuropsychological screens for CYP

Database searched: Scopus

Date of search: 05-11-2016

Search 1

| Step | Result |
|--|---------------|
| 1. Search: “Neuropsychological” (abstract, title, keywords) AND “screen” (abstract, title, keywords) Exclude: editorials and letters | 1205 articles |
| 2. Exclude: Articles from agri-sciences | 1181 articles |
| 3. Exclude: “Aged”, “Elderly” and “Middle Aged” (keyword) | 330 articles |
| 4. Exclude: Non-english language articles | 271 articles |
| 5. Exclude: Engineering articles | 265 articles |
| 6. Exclude: animal study (keyword) | 247 articles |
| 7. Reviewed abstracts Exclude: articles focussing on adults | 72 articles |
| 8. Reviewed abstracts Exclude: articles focussing on ASD | 53 articles |
| 9. Reviewed abstracts Exclude: clinical trials of interventions | 36 articles |
| 10. Reviewed abstracts Exclude: Screens of developmental conditions, e.g. FASD and childhood mental health difficulties | 13 articles |
| 11. Reviewed abstracts Exclude: articles focussing on infants, non-cognitive screens and projective tests | 7 articles |

Search 2

| Step | Result |
|---|---------------|
| 1. Search: “Cognitive” (abstract, title, keywords) AND “screen” (abstract, title, keywords) Exclude: editorials and letters | 4235 articles |
| 2. Include: Child OR Adolescent (keyword) | 508 articles |
| 3. Exclude: RCT or clinical trial (keyword) | 441 articles |
| 4. Exclude: Non-english language articles | 419 articles |
| 5. Exclude: Articles from outside medicine, psychology and social sciences | 397 articles |
| 6. Exclude: “older adults”, “aged” and “middle aged” (keywords) | 168 articles |
| 7. Reviewed titles Exclude: articles focussing on use of computer screens | 159 articles |
| 8. Reviewed abstracts Exclude: screens for ASD | 140 articles |
| 9. Reviewed abstracts Exclude: screens for mental health | 120 articles |
| 10. Reviewed abstracts Exclude: Experimental studies relating to cognition | 24 articles |
| 11. Reviewed abstracts Exclude: Screens for other developmental disorders, e.g. FASD and learning disability | 3 articles |

Overall result: 7 articles + 3 articles, with 2 articles overlapping = **8 articles**

Appendix B: Systematic literature review of use of games as neuropsychological / cognitive tests

Database searched: Scopus

Date of search: 05-11-2016

Search 1

| Step | Result |
|--|-----------|
| 3. Search: “Game” (Abstract, Title, Keyword) AND “Assessment” (Abstract, Title, Keyword) AND (“Cognition” (Abstract, Title, Keyword) OR “Neuropsychological” (Abstract, Title, Keyword)) Exclude: results from outside medical, psychological and social sciences fields and editorials | 86 papers |
| 4. Reviewed titles, identified common themes Exclude: Papers relating to game-training for rehabilitation, game theory in experimental psychology, game strategies in experimental psychology. | 13 papers |
| 3. Reviewed abstracts, identified wide range of topics not relevant to project, e.g. machine learning | 0 papers |

Search 2

| Step | Result |
|--|-----------|
| 1. Search: “Play” (Abstract, Title, Keyword) AND “Assessment” (Abstract, Title, Keyword) AND (“Cognition” (Abstract, Title, Keyword) OR “Neuropsychological” (Abstract, Title, Keyword)) Exclude: results from outside medical, psychological and social sciences fields and editorials | 65 papers |
| 2. Reviewed titles, identified common themes. Exclude: Papers relating to sports rehabilitation, computer game playing and cognition, play difficulties as symptoms of disorder, play therapy and play and longitudinal outcomes. | 8 papers |
| 3. Reviewed abstracts, Exclude: papers only relevant to assessment of infants. | 2 papers |
| 4. Reviewed abstracts. Exclude: paper relating to artificial intelligence | 1 paper |
| 5. Reviewed abstract. Exclude: remaining paper relating to effects of computer game play on cognition. | 0 papers |

Appendix C: Final set of faces used in Guess Who (not to scale)



Appendix D: Distribution of traits within the 24 Guess Who faces

| Trait | Female | Grey Hair | Ginger Hair | Blonde Hair |
|-------------------------------|-------------------|-------------------|-----------------------|-------------------------|
| Present in/ not present in | 11/13 | 4/20 | 2/22 | 4/20 |
| Black Hair | Brown Hair | Short Hair | Straight Hair | Bow/hat in hair |
| 7/17 | 7/17 | 16/8 | 17/7 | 18/6 |
| Hat | Bow | Bald | Fat/round face | Thin/narrow face |
| 4/20 | 2/22 | 4/20 | 4/20 | 4/20 |
| Long chin | Blue Eyes | Glasses | Freckles | Small Mouth/lips |
| 3/21 | 9/15 | 6/18 | 1/23 | 9/15 |
| Big mouth/lips | Big nose | Beard | Moustache | Smiling/Happy |
| 7/17 | 7/17 | 2/22 | 2/22 | 13/11 |
| Sad | Black | Earrings | | |
| 3/21 | 5/19 | 5/19 | | |

Appendix E: Pre-testing screen

- Could you count the dots in each of the squares here, without pointing with your finger? (Present dot-counting from ACE-III)
- Could you make up a sentence that contains a noun (a thing) and a verb (an action word).
- Could you take this piece of paper in your right hand, fold it in half and place it on the floor).
- Could you repeat these numbers in the order I say them: 2 1 8 5 4

Now, could you repeat these numbers in reverse order: 7 4 2

Appendix F: Guess Who Instructions

Experimenter says:

“Hi, I’d like you to play this game with me, it’s called Guess Who. It’s a game you might enjoy playing but I will also take some brief notes as we go along. I will be doing this as I hope that we can use the game to measure your thinking skills, such as your memory and attention. I will also be timing each game. However, you do not have to play quickly, it is more important to play the game well than quickly. Your job is to try to find out which face I have in my hand by asking me questions about it. Try to find out which face I am holding as quickly as you can. Here we’ve got some faces on this board (point to the board). You can flip the faces up or down (demonstrate, then flip all faces up). As you can see, the faces are all slightly different, can you tell me something about this person’s face?”

Show picture of Roger, see if examinee can describe something about Roger. If the participant fails to describe Roger’s appearance, offer a prompt that Roger has a beard. If the participant makes an overly subjective or speculative description, e.g. that he seems to be a nice guy, say

“Well, we can’t be sure about that, but we can be sure that Roger has a beard.”

Say:

An important rule is that you can only ask me a question that I can give a yes/no answer to. For example, you can ask me if they have brown eyes, but you cannot ask me “Do they have long hair or short hair?”, as I cannot answer yes or no to that (Ask examinee to repeat rules, check for comprehension.) I’m going to pull one of these faces out of this pack here (show deck of cards to examinee). It’s going to be one of the faces on the board here. I’d like you to ask me some questions about the face I have in my hand to try to find out which one it is, okay? You will pull out a face from the pack as well and I will try to find out who it is. The winner will be whoever gets the answer first. Okay, let’s start. I’m going to let you go first on each game.” During game, provide verbal encouragement regardless of performance. If the examiner wins a game, continue to play the game with the examinee until completion. Play the game 5 times, allowing the examinee to take the first turn on the first game. On the grid sheet, record the face turned down by the participant on each question (i.e. if a face was turned down after question 2, put “2” on this face’s position on the grid). Record the question and replies on the answer sheet. Discontinue at 20 questions, saying: *Okay, shall we start again with a different face?*

Prompts

If an examinee's first question for an item refers only to one object (e.g., "Is it Anna?"), record and answer the question. Then say, *“Remember, try to ask the*

fewest number of questions you can." Provide this prompt only once for each game.

When answering questions, respond only with yes or no as much as possible. In deciding how to answer, base your response on how most people would respond to the same question. If the question could possibly be answered either way, you may say, *Most people would say yes* or *Most people would say no*.

If the participant asks a question that is subjective, vague or tangential, record this as a question asked and say *It's difficult for me to tell from looking at the picture, could you ask me a question regarding their appearance?*. Record the next question as a separate question. Do not provide the prompt again during the game that is underway.

If an examinee asks a compound question (e.g., "Is it red and/or a plant?"), record the response and say, *I can answer only one of those questions using yes/no. Which one do you want me to answer?* If the examinee provides a yes/no question that clarifies the compound or either/or question, consider both questions as representing one yes/no question. If they ask a different question, record it as a separate question.

If an examinee fails to identify the target character after 20 questions but wants to know which one it is, say, *I can't tell you, but try to guess the next one*.

Once the final game is finished, put all of the faces down on the participant's board, say:

Now I am going to show you 6 faces that you will have seen on the board. I would like you to point to where you think you saw the faces on the board. Do not turn the faces over to check where they were even after you have pointed to where you think they are. If you're not sure of where a face was, try to guess as best you can.

Afterwards say:

Try to keep those in mind, I will ask you again in a little while where those same faces were again.

Repeat the above instructions after 20 minutes.

Appendix G: Response sheet for Guess Who

Participant ID: _____ Date: _____ Colour: _____ Person on card: _____

Order of Elimination

| | | | | | |
|--|--|--|--|--|--|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Questions

| No. | Question | Strategy Score | Error |
|-----|----------|----------------|-------|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | |
| 20 | | | |

Strategy / Abstraction

Strategy Score / question: _____

Pseudo-constraint Q's: _____

Tangential/Irrelevant: _____

Subjective: _____

Repeated questions: _____

Compound questions : _____

Spatial questions: _____

Turn-taking

Errors made: _____

Examiner Q's: Hat? Blonde Hair? Female?

Comments: any evidences of impulsivity, perseveration, perceptual difficulties, attention etc.

Appendix H: Order of questions for guess who.

| | | | |
|---------------|--------------------|---------------------|----------------------|
| Game 1 | Hat? | Blonde Hair? | Female? |
| Game 2 | Blue Eyes? | Earrings? | Male? |
| Game 3 | Female? | Black? | Happy? |
| Game 4 | Glasses? | Male? | Q about hair? |
| Game 5 | Brown Hair? | Male? | Hat? |

Appendix I: Recording Form for Guess Who Immediate and Delayed Recall.

Memory Task

| Immediate | | | | | |
|-----------|--|--|--|--|--|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

| Delayed | | | | | |
|---------|--|--|--|--|--|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Immediate Score: ___ / 24

Delayed Score: ___ / 24

Appendix J: Instructions for Connect Four

“So, now I’d like you to play this game with me, it’s called Connect Four. Like when we played Guess Who, I will take some brief notes as we go along and I will be timing us as well. However, you do not have to play quickly, it is more important to play the game well than quickly.”

The aim of the game is to be the first person to line up four of these tokens (show) in a row. A line can go up and down the board (show), across the board (show), or diagonally like this (show). We will take every second turn. You went first in Guess Who, so I will go first on each of the five games here. Would you like to be red or yellow?

Remember, you are trying to build a line of four tokens and so am I. The winner will be the person who builds the line first, so it is important to try to build your own line while at the same time stopping me from building mine. Shall we start?

Following each game, verbal encouragement was provided regardless of performance.

Appendix K: Guess Who recording sheet

| | | | | | | |
|---|---|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Switches: _____

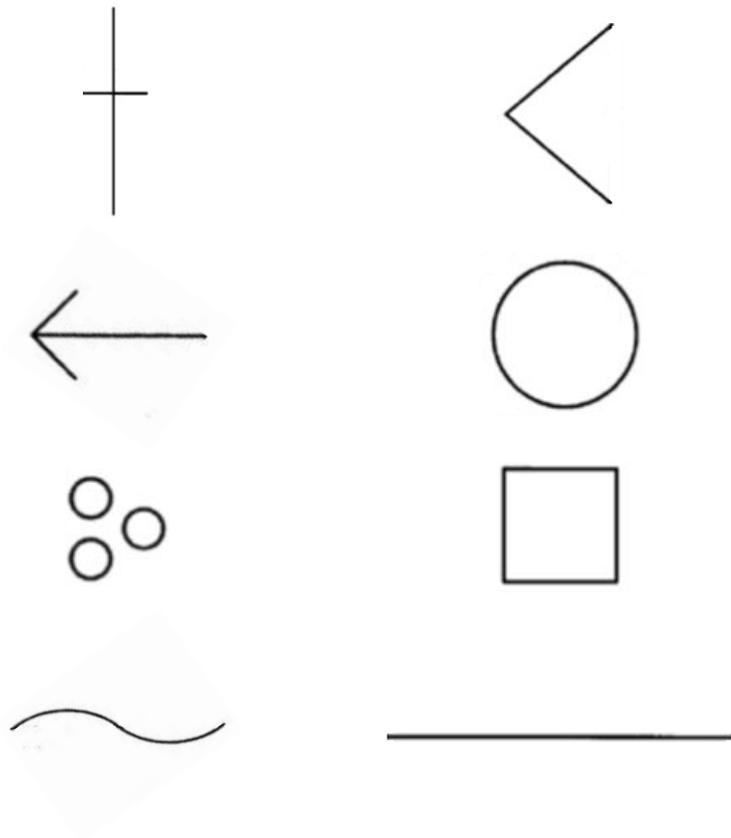
Missed winners: _____

Total switches/moves: _____

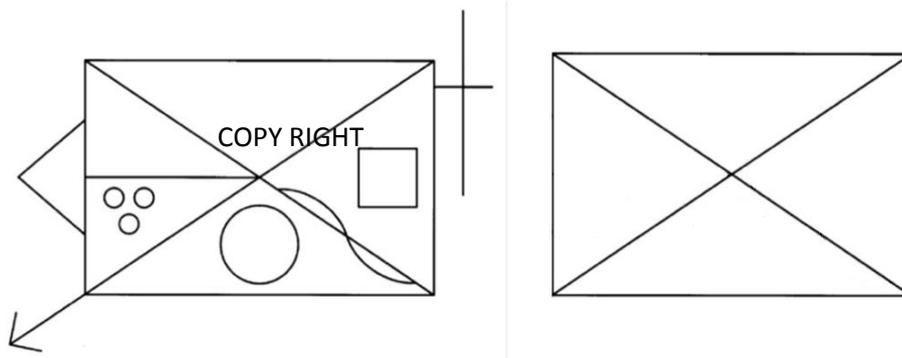
Identified winners: _____

Identified winners/total winners: _____

Appendix L: Eight missing RBANS components, complete figure and outline presented to participants at recall stage



Eight Missing Components



Complete Figure

Figure Outline

Appendix M: RBANS scoring grid

| Item | Placement | Orientation | Size |
|-----------------|------------------|--------------------|-------------|
| Cross | | | |
| Chevron | | | |
| Arrow | | | |
| Circle | | | |
| Circles | | | |
| Square | | | |
| Wave | | | |
| Line | | | |
| Total/25 | | | |

Appendix N: Engagement measures

How fun was completing the previous tasks for you?

Really boring,
I would not
like to do it
again.



1

Not much fun

2

Okay



3

A little fun

4

Really fun, I
would like to
do it again.



5

Did you feel anxious or worried while completing the previous tasks?

No, not at all.



1

Yes, just a
little

2

Yes, quite a
bit

3

Yes, I felt
really anxious,
I almost
wanted to stop



4

Appendix O: Information Sheet for Adults

Project Title

Using board games as neuropsychological tests with children with acquired brain injury.

The Chief Researcher

Patrick Murphy

Email: u1438315@uel.ac.uk

Telephone: 07757 218 742

Project Sponsor

University of East London.

Invitation to participate in a research study.

This letter provides parents or guardians of children and young people with information regarding a research project they have been invited to participate in. This can allow you help them to understand the research project and decide whether they should participate in the study.

The study is being conducted as part of my Professional Doctorate in Clinical Psychology at the University of East London. The project has received approval from the University of East London and NHS Research Ethics Committees.

Description of the project.

When children and young people experience a brain injury, they commonly complete neuropsychological tests with a psychologist. These tests involve trying to remember and recall information, using thinking skills and recognizing objects. They allow us to understand how brain functions such as memory and attention have been affected. They can also help medical staff understand how the brain has been affected by injury.

In this piece of research, we interested to find out whether board games such as Guess Who? and Connect 4 can be used as neuropsychological tests, i.e. to test memory, attention, and other thinking skills. We are also interested in whether these games are more engaging than traditional neuropsychological tests. Finally, we are interested in whether these games are helpful in predicting everyday difficulties following brain injury.

We are approaching both children and young people with and without a brain injury to participate in this project.

Why have I, as a parent/guardian, been approached?

By law, children and young people under the age of 15 (either with or without a brain injury) require the consent of their parents before taking part in any research. As a parent or guardian of a (prospective) participant in the study, you will have an interest in your child's well-being and welfare and can make a decision as to whether it is in their interest to participate. You are also in

a position to understand any views they may have about taking part in such a project, but may have difficulty communicating. Finally, you may be able to tell us about any possible difficulties they may have if they participate in the research, i.e. how they might indicate if they become distressed and want to stop, and communicate this information to us.

Any children or young people who can understand the research in general, i.e. the activities involved and risks and benefits, will also be asked to indicate their willingness to participate by signing a form.

What your child will do.

Your child will play Guess Who? and Connect 4 (maximum 6 games each) with the researcher. This should take no longer than 40 minutes. Following a break of approximately 20 minutes, he/she will then complete some standard neuropsychological tests, which will take approximately 60-90 minutes.

Parents will be asked to complete a short questionnaire about aspects of the child or young person's thinking skills, such as their attention and memory.

Other information sought for this research.

We are interested in gathering some background information on your child as well. Therefore, we will ask some basic questions about educational level, occupation and cultural background of you and your child. Where possible, we would like your help when gathering this information.

Potential benefits of the research.

The primary goal of this research is to provide a more engaging, more "everyday" and less distressing method of conducting neuropsychological testing with children and younger people. We hope to publish the findings of our research in scientific journals so our work can benefit the wider population of children and younger people with brain injury. If you would like feedback on the overall findings of our project, please contact either me or my supervisor (details below) and a summary of the findings will be sent at the end of the project.

The results of the tests carried out in this research may be helpful for your child's education. Therefore, we can provide a brief report of the results of the tests if so desired. The implications of these results can be discussed with educational staff.

Is this part of my child's education or (if applicable) care or rehabilitation?

No. This research is not part of your child's education. Anyone who decides to not participate in this research will do so at no disadvantage to themselves.

The right to withdraw.

Your child is free to withdraw from the study at any time at no disadvantage to himself/herself.

All data will be destroyed two years following completion of the research (completion date: May 2017) or following publication of the data, whichever date is earlier. At this point, only aggregated data for all participants will remain and withdrawal will not be practically possible.

Risks of participation.

Completing these tests can sometimes be stressful. Therefore, your child will be reminded that they are free to withdraw from the study if they become visibly anxious or distressed. We do not see any additional risks to participants beyond this.

The researcher has passed all appropriate Disclosure and Barring Checks for working with children.

What if participants would like feedback on their results?

As mentioned above, a report on the results from the tests can be provided if desired. These can be discussed with education or healthcare staff for further information on the implications of the results.

Right of children and young people to decide to participate

It will be assumed that your child can make the decision themselves whether to participate or not. However, we will check before beginning any research tasks that they understand what is involved in the research and can weigh up to pros and cons of participating.

Confidentiality of the Data

All data gathered will be retained in accordance with the University of East London's Data Protection Policy. All data will be kept completely confidential and only available to the researcher and his supervisor. All paper records of collected data will be kept in a locked file for research material at the University of London. Initials rather than full names will be used on answer sheets to ensure confidentiality.

Once gathered, all data will be transferred to a database for analysis on an encrypted UBS key. However, personal information will not be entered into this database. Your child will be identified with a number in the database to ensure that they cannot be identified. A separate file will provide a key for your child's ID number, which will be kept on a separate piece of hardware. All electronic files used for handling your child's data will be password protected.

Paper records of gathered data will be destroyed immediately following completion of the research. As described above, all other data will be destroyed two years following completion of the research, or following publication of the research. Electronic files will be permanently deleted and paper records shredded.

Confidentiality may be breached without you or your child's knowledge if they disclose that either they or someone else are at serious risk of harm.

Location

The research will take place at school, at the University of East London, or the participants' residence, whichever is most convenient.

Remuneration

Unfortunately, it is not possible to provide payment for participation in this research. In some cases, it may be possible to provide a contribution towards vouched travel expenses and lunch for participants. A light snack will be provided for all children and parents that participate.

Disclaimer

Your child is not obliged to take part in this study. He/she should not feel coerced into participation and is free to withdraw at any time. Should he/she choose to withdraw from the study they may do so without disadvantage to themselves and without any obligation to give a reason.

Please retain this invitation letter for reference and feel free to ask me any questions.

If you have any further questions or concerns about how the study has been conducted, please contact me, the principal investigator, at the email address or phone number above.

Alternatively, please contact the study's supervisor:

Dr. Jenny Jim,
School of Psychology,
University of East London,
Water Lane,
London E15 4LZ.
Telephone: 020 8223 4411
Email address: j.jim@uel.ac.uk

Alternatively, if you have any concerns about the conduct of the investigator, researcher(s) or any other aspect of this research project, you can contact the following with your concerns

Catherine Fieulleateau,
Research Integrity and Ethics Manager,
The Graduate School, Docklands Campus,
University of East London, London,
E16 2RD
Telephone: 0208 223 6683
Email: researchethics@uel.ac.uk

Thank you in anticipation.

Yours sincerely,

Patrick Murphy
Chief Researcher

Appendix P: Information Sheet for CYP

University of East London
School of Psychology
Stratford Campus
Water Lane
London E15 4LZ



Invitation to take part in a research project: part 1

Hi, my name is Patrick Murphy. I am studying to be a psychologist at the University of East London. I am inviting you to take part in a research project I am running. The following sheet will tell you what the project is about.

What is a research project?

A research project is about finding out new facts. It's different to, say, looking for the answer on the internet, as it is about finding out something new that hasn't been looked at before.

A scientist might do research. She might use an experiment to find out something new about the world around us.

A doctor might also do research. He might provide people with a new medicine and afterwards see if it helped them feel better.

Once research is completed, the results are shared with other people. Hopefully, others will benefit from what the research has found out.

What is this research project about?

We are interested in how children and young people are *assessed* after having a brain injury.

If a child or young person has a brain injury, they will complete *psychological tests*. These tests help us understand how the brain was affected by the injury. They also help us understand whether thinking skills have been affected by the brain injury.

In this research project, we would like to see whether the games Guess Who? and Connect 4 can be used as psychological tests. We are looking for young people who have had a brain injury and who have not had a brain injury to take part.

What are you interested in this?

There are two main reasons we are interested in this.

Firstly, many children and younger people with a brain injury find psychological tests boring or even upsetting. If they were able to play a game instead of the tests, this might make things easier and more enjoyable.

Secondly, by playing a game, the psychologist will get to know the child or young person much better than if they just watch them do a test. This helps a psychologist better understand any difficulties the child or young person might have.

Who might this help?

We hope that this research will help children and young people who have a brain injury. We hope that this research might make assessment less stressful for them.

We hope that this research will also help psychologists, therapists, doctors and nurses who work with children and young people who have a brain injury. We hope it will help them complete their assessments.

If you decide to take part in the research we can write a report for you on how you did on the tasks. This might be helpful for your school. If you are working with a therapist it might help them also.

If this sounds interesting to you, please read the other information sheet in this envelope. It will explain the tasks in the research to you.

Information on a research project: part 2

The first sheet told you why we are doing this research. This sheet tells you what you will do if you decide to take part. We hope the information in this sheet will help you decide whether you would like to take part or not.

What tasks will I have to do?

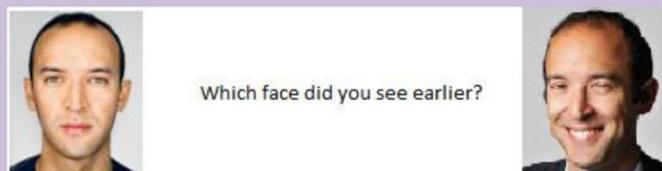
In the first part of the research, you will play games of Guess Who? and Connect 4 with the researcher. You will play each game 6 times with the researcher.

In Guess Who?, both players ask each other questions to try to find a mystery face. Whoever guesses the mystery face first wins the game.



In Connect 4, both players try to make a line with 4 discs in a rack. One player uses yellow discs and the other red discs. They take every second turn. Whoever gets 4 in a line first wins.

After this, you will complete some psychological tests. These tests will look at thinking skills and your memory. For example, you will look at some faces and try to recognise them afterwards.



At the end of the research, you will be asked some questions about how you enjoyed the different tasks.

How long does this take?

It will take about 40 minutes to play the games and 60 to 90 minutes to complete the psychological tests. There will be a 20 minute break between these. So it will take about two and a half hours to complete everything.

Who will find out about my results?

We keep your information *confidential*. Only people involved in the research can see it. We do not write your name on any of the sheets we use. We do not write your name on the computer files we keep. This makes sure that no one else can see your results.

The only time we might share information about you without your permission is if we felt your safety was at risk.

We will make sure that your name is deleted from any records we have once the study ends (May 2017). All of the information we gather will be deleted by December 2018.



I would know how I did on the test, can you let me know?

Yes, if you would like a report explaining how you performed on the tasks, we can write one for you.

Do you tell my teacher or doctor how I did?

No. This is not part of your work at school or with anyone in the health service. We will not share any information with them without your permission.

Will you need any other information from me?

We would like one of your parents to answer some questions about you. They will be asked some questions about your thinking skills and behavior.

Do I have to take part?

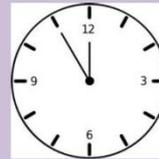
No. It is up to you whether you want to take part or not. You can stop taking part while we are playing the games or doing the psychological tests. If you do not want to take part afterwards, you can tell us to delete your results from our files.

I'm worried that this could be difficult ...

Sometimes, people can become worried or upset while doing research like this. If this happens you can ask to take a break. You can also stop at any time if you want.

Do I have to give up my own time to do this?

Yes. We are not able to do the research during school time.
If you would like to take part, we will try to meet you at a good time for you.
We cannot pay anyone for taking part in the research.



Where will it happen?

We can complete the tasks either at your home, or at the University of East London in Stratford, East London.

I have a few questions, who should I talk to?

Please contact the chief researcher, Patrick Murphy. His contact details are at the end of this sheet. You can also discuss the research with your parents.

I'm interested in taking part, what should I do? What will happen next?

If you can, please talk to your parents first. After this, you can telephone or email Patrick Murphy, the chief researcher. He will arrange a time to meet with you and do the research tasks.

Before the research begins, you will complete a few short forms. These are to show that you understand what you are going to do. If you do not understand everything in the research, we will need to get your parent's permission also before we start.

Thank you for reading this sheet,

Patrick Murphy,
Chief Researcher
07757 218742
U1438315@uel.ac.uk

Dr. Jenny Jim
Research Supervisor
jjim@uel.ac.uk

Appendix Q: Consent form for parents

UNIVERSITY OF EAST LONDON

Consent to participate in a research study (parent/guardian version 1.0)

Using board games as neuropsychological tests with children with acquired brain injury.

- Please
circle
1. I have read the information sheet (Version 1.0) relating to the above research study and have been given a copy to keep. The nature and purpose of the research have been explained to me. I have had the opportunity to discuss the details and ask questions. I understand what is being proposed and the procedures involved have been explained to me. Yes / No

 2. I understand that the participant's involvement in this study, and particular data from this research, will remain strictly confidential. Only the researcher(s) involved in the study will have access to identifying data. It has been explained to me what will happen once the research study has been completed. Yes / No

 3. I understand the purpose of the project and what the participant's involvement would be. In my opinion, they would not object to taking part in the study. Yes / No

 4. I understand that participation in the project is voluntary and that the participant would be withdrawn if they do not wish to continue participating and without giving a reason with no further disadvantage to themselves. Yes / No

 5. I understand that data from this project will be presented at conferences and submitted for publication in journals and I consent to my data being used in this way. I understand that my data will be in no way identifiable when disseminated. Yes / No

 6. I agree to participate in the study by completing a proxy report of the participant's difficulties. I understand what this involves and that the confidentiality of the data and my right to withdraw or cease participation are as described in 2 and 4 above. Yes / No

Parent / Guardian's Name (BLOCK CAPITALS)

Parent / Guardian's Signature

Researcher's Name (BLOCK CAPITALS)

Researcher's Signature

Date:

Appendix R: Assent form for CYP

UNIVERSITY OF EAST LONDON

Assent form for children and young people

Using board games as psychological tests with children with brain injury.

Please circle

I have read the information sheet on the research. I have been given a copy of the information to keep. Yes / No

The research has been explained to me. I have been able to ask questions about the research. I understand what I will have to do if I take part. Yes / No

I understand that information about me will remain confidential. Only the researchers involved in the study will be able to see my information. I understand what will happen to my information once the study has been completed. Yes / No

I understand that if the researcher becomes concerned for my safety or the safety of anyone else during the research he/she may have to inform someone else (for example, the NHS or the police) without my permission. Yes / No

I agree to take part in the study. I understand that I can stop taking part in the study at any time without having to explain why. Yes / No

I understand that this research is not part of my education or healthcare. I understand that if I do not take part, or stop taking part, this will not affect me in any way. Yes / No

Participant's Name (BLOCK CAPITALS)

Participant's Signature

Researcher's Name (BLOCK CAPITALS)

Researcher's Signature

Date:

