

USING A GAME-LIKE PROCEDURE  
AS A NEW TEST OF PROBLEM  
SOLVING AND CONCEPT  
FORMATION IN CHILDREN.

**Annie Pavitt**

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

## ABSTRACT

Concept formation is a complex cognitive process which allows us to identify, categorise, and determine rules from sets of experiences alongside developing cognitive representations for complex thoughts, behaviours, and events. Concept formation abilities are highly important for children's performances in school, socially, and therefore throughout their lives. Yet, most of the tests that are used to measure concept formation in children were developed for use with adults; very few have been specifically designed for children.

This study is the first phase of test development. It addressed whether a new game of concept formation, *the Alien Game*, could capture executive functions in a group of 18 typically developing children. The game was based on the 20-Questions task and a game-like structure was adopted in order to engage participants. Codes were generated from a two-phase content analysis carried out using previous research and the children's responses within this study. The task was also interpreted using an Abstraction Score, a Learning Slope, an Initial Abstraction Score, a Weighted Achievement Score, and a Time to First Question measure.

The content analysis indicated that similar patterns were found among the children's responses, with the children using a majority of 'Constraint seeking' questions. The Abstraction Score, Initial Abstraction Score, and Weighted Achievement Score addressed separate aspects of the questions asked, and were each designed to assess the quality of questions asked.

These results indicate that this task could potentially be used as a formal cognitive test. This needs to be further developed through future research on norms, reliability, and validity. Clinically, the Alien Game shows potential of being used as a neuropsychological tool which could be used to capture deficits in children's concept formation early, in order for appropriate interventions and support to be put in place at a younger age.

## ACKNOWLEDGEMENTS

I would firstly like to thank my participants for giving up their time to participate in my study and for your enthusiasm. I learnt a lot from you and the different ways you approached this task. I would also like to say a big thank you to my teacher friends for arranging for me to come into their schools and to recruit some of their class into my study. This included sending out information sheets to their pupil's parents and answering their questions – I am hugely grateful for all of your time and effort spent! I would also like to thank the Head Teachers and SENCOs at these schools for allowing me to recruit children into the study.

I would next like to thank Dr Matthew Jones-Chesters for all of his guidance, wisdom, and support throughout both the development of my study and in the write up of my thesis. I have learnt a lot from his useful insights and am very grateful for all of his encouragement and responsiveness throughout this process.

A special thank you goes both to my dad and to Lawrence. Thank you both for being there and offering an abundance of support, both emotional and practical. Your relentless belief in me has encouraged me to keep on going at times when it felt difficult. I would also like to thank the rest of my family: my mum, siblings and their partners, Yvonne, and granny. I am not sure how to put into words how grateful I am to have such a supportive, caring, and fun family around me.

Finally, a huge thank you goes to the rest of my friends for sticking by me and being so understanding and supportive during the doctorate course which perhaps seemed to be never-ending at times. I feel incredibly lucky to have such supportive and caring friends. I am also grateful to have met and shared this experience alongside my cohort. Being part of such a supportive year group has made these last three years a lot of fun alongside the hard work.

## TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES AND FIGURES .....	vi
1. INTRODUCTION .....	1
1.1. Executive Function.....	2
1.2. Concept Formation .....	3
1.3. Literature Review .....	4
1.3.1. Methods of Literature Review .....	4
1.3.2. Inclusion and Exclusion Criteria .....	4
1.3.3. Search Results .....	4
1.4. Review of Literature.....	5
1.4.1. Concept Formation in Children.....	6
1.4.2. Assessment of Concept Formation.....	16
1.4.3. Aetiology of Difficulties.....	23
1.4.4. Impact of Deficits .....	26
1.5. Summary of Literature Review.....	29
1.6. Present Study .....	30
1.6.1. Aims and Rationale .....	30
1.6.2. A Game-Like Procedure .....	31
1.7. Research Questions .....	32
2. EPISTEMOLOGY & METHODOLOGY.....	33
2.1. Epistemological Approaches.....	33
2.2. Adopting a Critical Realist Approach .....	34
2.3. Inclusion & Exclusion Criteria.....	35
2.4. Design.....	35
2.5. Materials .....	36
2.6. Ethics .....	38

2.6.1. Confidentiality .....	39
2.7. Procedure .....	39
2.8. Data Collection and Analysis .....	40
2.8.1. Content Analysis .....	40
2.8.2. Analysis of the Coding Data .....	43
2.8.3. Additional Tests .....	43
2.8.4. Acceptability .....	44
2.9. Participants .....	45
2.9.1. Demographic Data .....	45
3. RESULTS .....	49
3.1. Methods of Analysis .....	49
3.2. Content Analysis .....	50
3.2.1. Quantitative Analysis .....	54
3.2.2. Inter-Rater Reliability .....	54
3.2.3. Exploratory Content Analysis .....	56
3.3. Other Measures .....	58
3.3.1. Abstraction Score .....	59
3.3.2. Learning Slope .....	60
3.3.3. Initial Abstraction Score .....	61
3.3.4. Weighted Achievement Score .....	62
3.3.5. Comparison with Other Measures .....	63
3.3.6. Time to First Question .....	65
3.3.7. Acceptability .....	66
4. DISCUSSION .....	67
4.1. Revisiting the Research Questions .....	67
4.1.1. Research Question 1 .....	67
4.1.2. Research Question 2 .....	69
4.1.3. Research Question 3 .....	73
4.2. Summary of Findings .....	73
4.3. Evaluating the Present Study in Relation to Previous Research .....	74
4.4. Critical Evaluation .....	79
4.4.1. Strengths of the Current Study .....	79
4.4.2. Limitations of the Current Study .....	81
4.4.3. Research Reflexivity .....	85

4.5. Conclusions .....	86
4.6. Clinical Implications .....	87
4.7. Future Research Directions.....	88
4.7.1. Reliability .....	89
4.7.2. Validity .....	90
4.8. Concluding Summary .....	91
 REFERENCES.....	 92
 APPENDICES: .....	 108
APPENDIX A: Pictures Used in the Alien Game. ....	109
APPENDIX B: Consent Form – Child. ....	109
APPENDIX C: Information Sheet .....	109
APPENDIX D: Opt Out Form for Parents/Guardians. ....	109
APPENDIX E: Consent Form – Parents/Guardians.....	109
APPENDIX F: Head Teacher’s <i>Loco Parentis</i> Form.....	109
APPENDIX G: Standardised Instructions for Child. ....	109
APPENDIX H: Measure of Task Acceptability Using Likert/Visual Analogue Scale ...	109
APPENDIX I: Coding Sheet .....	109
APPENDIX J: Coding Scheme Explained.....	109
APPENDIX K: Scoring for Code 7 and Code 8.....	109
APPENDIX L: Ethical Approval .....	109
APPENDIX M: Mean Learning Slope Total and in Different Age Groups.....	109

## LIST OF TABLES AND FIGURES

<b>TABLE 1.</b> Developmental Trajectory of Concept Formation.....	11
<b>TABLE 2.</b> Codes From Previous Tests of Concept Formation.....	42
<b>TABLE 3.</b> Non-Parametric Correlations Between and Within National Curriculum Levels for Reading, Writing, and Maths.....	46
<b>TABLE 4.</b> Descriptive Statistics for Age and Key Cognitive Tests.....	47
<b>TABLE 5.</b> Measures of Association Between the Demographic and Cognitive Scores.....	48
<b>TABLE 6.</b> Two Phase Content Analysis.....	51
<b>TABLE 7.</b> Coding Scheme.....	53
<b>TABLE 8.</b> Percentage of Codes Used and Number of Questions Asked Overall.....	55
<b>TABLE 9.</b> Mean Abstraction Scores.....	59
<b>TABLE 10.</b> Mean Learning Slope Scores.....	60
<b>TABLE 11.</b> Initial Abstraction Score.....	61
<b>TABLE 12.</b> Weighted Achievement Scoring System.....	62
<b>TABLE 13.</b> Weighted Achievement Score.....	63
<b>TABLE 14.</b> Spearman’s Rank Correlation Comparing Abstraction Score with Similarities and Matrix Reasoning.....	64
<b>TABLE 15.</b> Friedman’s Two-Way Analysis By Ranks.....	66
<b>TABLE 16.</b> Coding Scheme Explained.....	119
<b>TABLE 17.</b> Scoring for Code 7 and Code 8.....	121
<b>TABLE 18.</b> Scoring for Code 7 and Code 8 – Continued.....	122
<b>TABLE 19.</b> Mean Learning Slope Total and in Different Age Groups.....	126
<b>FIGURE 1.</b> PRISMA (2009) Flow Diagram of Article Selection Process.....	5
<b>FIGURE 2.</b> Related-Samples Friedman’s Two-Way Analysis of Variance by Ranks.....	66
<b>FIGURE 3.</b> Pictures of Aliens Created For the Alien Game.....	109
<b>FIGURE 4.</b> Example of a Coding Sheet Used by the Second Rater.....	118

## 1. INTRODUCTION

This research is situated within the field of assessment of concept formation in children. Gelman (1999) reported that there had been over 7,000 research articles written about children's concepts over the past thirty years; a number that will have since risen. Researchers are attracted to this area for the opportunity to explore fundamental issues such as how early childhood thinking is characterised, alongside practical concerns about how children form concepts which are directly pertinent to their abilities and performance at school and throughout their lives (Gelman, 1999). Yet despite the interest in and importance of this topic, there have been relatively few standardised neuropsychological tests designed to specifically test concept formation in children (Channon & Crawford, 1999). The aim of this present study is to examine the utility of using a game-like procedure in order to test children's concept formation and problem solving abilities.

This section will introduce the reader to the research topic. It will firstly position concept formation within the wider topic of executive functioning. It will then introduce the literature on concept formation in children, including a discussion on how concepts are acquired and the developmental trajectory of these abilities. This section will also discuss current and past neuropsychological tests used to assess concept formation in both children and adults. Throughout this section, a critical review of the literature is provided. The aetiology of difficulties with concept formation and the possible academic and psychosocial sequelae of having such difficulties will also be explored. Finally, the present study will be outlined as derived from the literature review, including a discussion of its rationale, aims, and research questions.

## 1.1. Executive Function

Executive function is a term used to describe a set of cognitive capacities intimately associated with the frontal lobes, specifically the prefrontal cortex (Morris, Ahmed, Syed, & Toone, 1993) and connections to other brain areas. There is currently no consensually agreed definition of executive function, however, it can be broadly defined as the capacity to engage in “independent, purposive, self-serving behavior” (Lezak, Howieson, & Loring, 2004, p. 42). Executive functions play a central role in helping people perform in new, non-routine, or complex situations that require integrating knowledge and experience (Welsh, 2002). It is a set of interrelated abilities responsible for enabling a person to manage, regulate, plan, and execute meaningful aim-oriented behaviour (Gligorović & Buha, 2013). Carlson, Zelazo, and Faja (2013) note that these skills may be observed in children when they pause and reflect before acting. Children with difficulties in executive functioning may find problem solving difficult as they may find it hard to plan, anticipate consequences, form concepts, and initiate activity towards goal-directed behaviour (Darby & Walsh, 2005). These children may also be slow to master reading skills (Lorsbach, Wilson, & Reimer, 1996).

Although an in-depth review of executive functioning is beyond the scope of this thesis, broadly speaking it can be broken down into three domains: working memory, goal orientated behaviour, and concept formation. These three aspects interact in typically developing people, although they are viewed in literature, and thus neuropsychological testing, as separate capacities with different functions. However, it is important to acknowledge that, as with any scientific construct, our current understanding of executive functions, and its domains, are situated within the present place and time.

## 1.2. Concept Formation

Although explanatory theories and specific definitions of concept formation differ somewhat, generally it is regarded as a complex and multifaceted cognitive process which allows people to derive concepts based on relational, functional, and perceptual aspects (Hills, Maouene, Maouene, Sheya, & Smith, 2009). This enables someone to categorize objects or experiences by sorting them under more general rules or classes (Hunt, 1962), for example, recognizing a chair without having seen that particular chair before. It enables us to identify, determine, and categorize rules from sets of patterns (McGrew, 2009) and develop cognitive representations for complex thoughts, behaviours, and events (Medin & Smith, 1984). Thus, concept formation is a key component of a person's information processing which forms the basis for many higher-order linguistic and cognitive processes including: abstraction, reasoning, controlled attention, and the ability to compare (Vygotsky, 1986).

The ability to generate and flexibly change concepts are crucial abilities in the development of academic and adaptive skills (Gligorović & Buha, 2013). Therefore, children with difficulties in these abilities can face challenges at school and in later life (Gelman, 1999). Despite this, few clinical tests specifically aim to assess concept formation in children (Channon & Crawford, 1999). It is not clear exactly why this is, but it could be based on the view that concept formation is underdeveloped in children (Starkey, 1979). In response to this, the researcher would now like to orientate the reader to an exploration of research on concept formation in children from a position of it being a measureable and important ability which develops throughout childhood. Thus, a new test of concept formation specifically aimed at children would be valuable to allow early identification and intervention for deficits in this important function.

### **1.3. Literature Review**

#### **1.3.1. Methods of Literature Review**

An exhaustive search of literature on concept formation in children was carried out using PsycINFO, CINAHL Plus, and PubMed databases. The reference lists from retrieved papers were also manually searched for relevant articles and publications.

#### **1.3.2. Inclusion and Exclusion Criteria**

The search was limited to empirical papers published in English in peer-reviewed journals within an unrestricted timeframe. Studies were included if they investigated concept formation in children. Only studies in which concept formation was directly measured or observed and reported were included. All study types were included in the review, including: empirical studies, clinical cases, longitudinal studies, and systematic reviews. The search included studies of the developmental trajectory of concept formation in typically developing children. Additionally, it included research which explored concept formation difficulties in children in order to explore the psycho-social effects of such deficits.

The search strategy aimed to identify all research in which concept formation in children was investigated alongside neuropsychological testing or the symptoms related to difficulties. The search terms were 'Concept Formation' along with Child\*, Psych\*, Neuropsych\*, and Test\* (where \* denotes truncated terms) which were used in various combinations.

#### **1.3.3. Search Results**

Figure 1. illustrates the number of papers included at each stage of the literature review process, following the PRISMA Flow Diagram guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). The initial search strategy identified 592 papers. After removing duplicates, the titles, and abstracts where necessary, were read to identify studies which broadly fit within the relevant area. Where titles and abstracts did not provide the required information, full papers were accessed to determine eligibility.

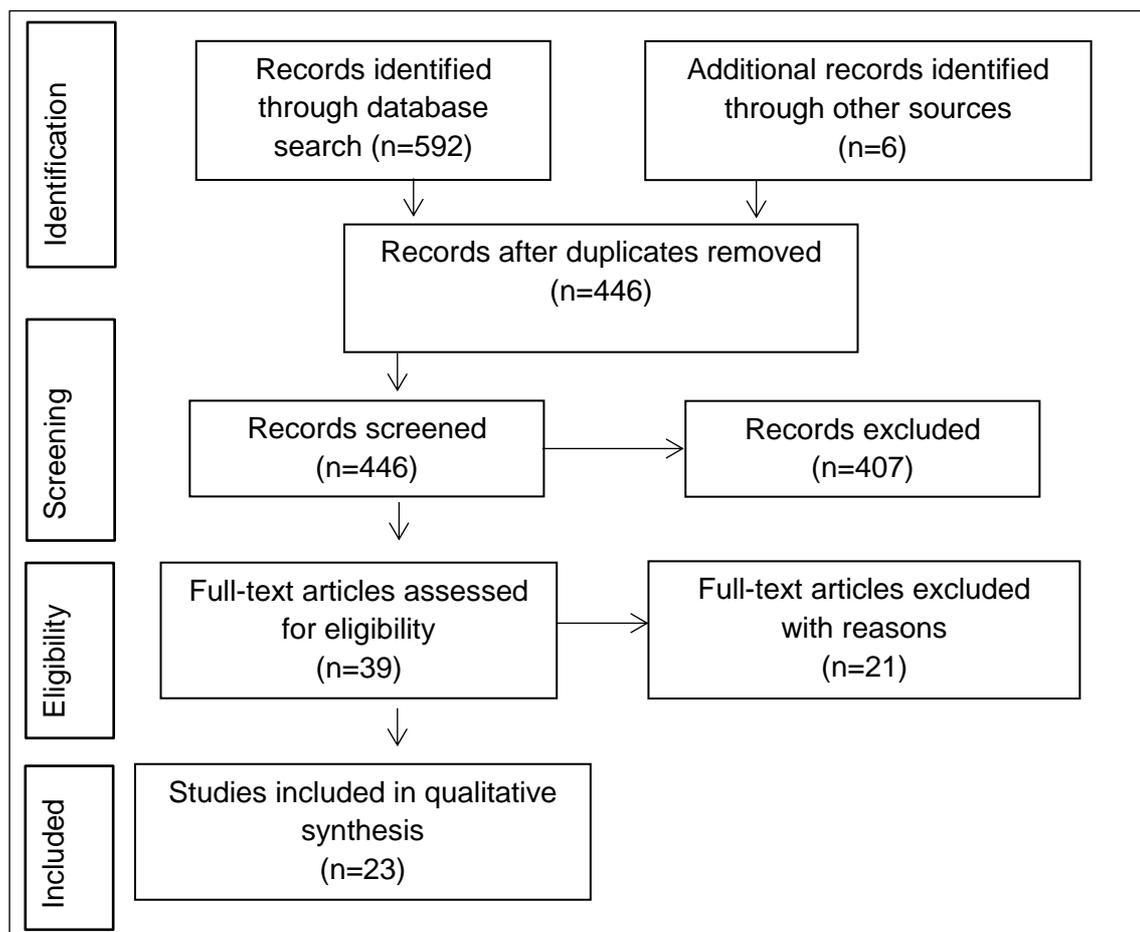


Figure 1. PRISMA (2009) Flow Diagram of Article Selection Process<sup>1</sup>

The next part of the report is derived from the qualitative synthesis of information obtained from the 23 papers remaining after this process.

#### 1.4. Review of Literature

In this section, I will start by reviewing contemporary literature on concept formation in children in order to define what concept formation is, and how and when typically developing children acquire these abilities. I shall also explore similarities and differences in how children and adults form concepts. I shall do

<sup>1</sup> From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred

Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097. doi:10.1371/journal.pmed1000097

this before reviewing how this literature has given rise to associated cognitive tests of concept formation which are currently available with a particular view to reviewing the strengths and weaknesses of these, alongside the supporting evidence and usefulness with children. The gaps in, and limitations of, these tests will also be discussed. I will then review literature on the aetiology of problems that children with concept formation difficulties may face in order to explore how difficulties in concept formation can manifest and thus why it is important to be able to test these abilities in children. After reviewing the literature, I will lead onto describing how this present study fits within gaps and recommendations from the contemporary literature reviewed in this sub-section with a view to creating a cognitive test that can best capture children's concept formation abilities as derived from the literature review.

#### 1.4.1. Concept Formation in Children

In order to assess the need for, and design of, a test of concept formation in children, this section will explore how and when children form concepts and whether children form concepts differently to adults.

1.4.1.1. *Definitions* - a key part of defining 'concept formation' is defining what is meant by the term 'concept'. There have been many different definitions which led Medin and Smith (1984) to describe the term as a 'loaded one', serving numerous explanatory functions within both psychology and related disciplines. This report will, however, adopt Mandler's (2008) definition of 'concepts' as knowledge about objects and events that are accessible to conscious thought.

Rey (1983) viewed concept formation as comprising four functions:

1. *Simple categorisation* – deciding if something can be categorised as part of a simple class (e.g. whether or not a particular object or being is an occurrence of the concept *girl*).
2. *Complex categorisation* – deciding if an object or being is an example of a complex class (e.g. concluding that an object is an example of the concept *rich girl*)

3. *Linguistic meaning* – explaining relations of semantic, synonymy, implication, and antinomy (e.g. the meaning of *lass* is somewhat synonymous to *girl* and indicates being young and female)
4. *Components of cognitive states* – including beliefs and preferences. In this function, concepts give a cognitive explanation for complexities of behaviour and thought (e.g. concepts *rich*, *girl*, and *spoiled* are played out in the resulting belief that *rich girls are spoilt*).

The categorisation process is fundamental to efficient information processing (Rosch, 1978). Simple categorisations provide the basis of more complex categories and have been a major focus throughout literature concerning itself with concepts (Smith, Jones, & Landau 1996). As this is the initial phase of test development, this study took an exploratory approach to assess which concept formation strategies the children appeared to adopt while problem solving.

1.4.1.2. *Approaches to concept acquisition* - one way children acquire new concepts is through the ability to detect new rules. There has been a long history of research in this area, including both visual and verbal concept acquisition. In Fantz's (1964) experiments of visual habituation, two stimuli were provided to infants; one was novel, the other stayed the same. He found that infants progressively looked towards the novel stimulus, indicating they had habituated to the familiar stimulus. These studies demonstrate infant's ability to acquire visual concepts. Other research focused on verbal concept acquisition. For example, McShane (1980) told children in his study that they were using an animal name incorrectly, and supplied them with the 'correct' name. He then found these children incorporated the new name into their vocabulary (e.g. "Thus Brian, told that what he called ducks were turkeys, called the toys turkey ducks ...", McShane, 1980, p. 112).

The ability to form concepts in novel situations starts early in a child's learning process, at a time when relatively few exemplars have been encountered (Mandler, 2008). However, forming concepts in novel situations does not stop in infancy; studies have found that, when given sufficient knowledge about an object or event, both children (Macario, Shipley, & Billman, 1990) and adults (Ahn,

Brewer, & Mooney, 1992) are able to accurately generalise a concept once exposed to a single instance. As the number of category examples increases, representations of categories are increasingly likely to incorporate both novel features and prior knowledge (Heit, 1994; Wisniewski & Medin, 1994). The ability to form categories based on prior knowledge of conceptual or perceptual differences and similarities allows novel stimuli to be organised in a manageable way (Alderson-Day & McGonigle-Chalmers, 2011). Therefore, exposure to new instances and novel features is an important part of concept formation which can refine and change previous knowledge or beliefs about categories.

Even young children do not just derive concepts solely on the basis of novel perceptual or sensory information (Keil, 1986; Mervis, 1987). Much research has suggested that from a very young age, children also interpret information with reference to existing domain knowledge and theories they hold about the world (e.g. Carey, 1985; Gelman & Markman, 1986; Keil, 1989). Krascum and Andrews (1998) found that when 4-5 year olds were able to use their pre-existing knowledge of animals this helped in their learning of new animal categories. Similarly, Carmichael and Hayes (2001) showed 4-10 year olds pictures of fictitious animals which were either consistent or inconsistent with the child's pre-existing knowledge and found that prior knowledge influenced their categorisation in all experiments. Barrett, Abdi, Murphy, and Gallagher (1993) also found that 5-10 year old children link specific features associated with individual concepts to form theories. Their experiments indicated that theory-based correlations were prominent in decisions regarding category membership and children's theories determined which attributes were considered most central to the concept. Furthermore, Farrar, Raney, and Boyer (1992) found younger children made more inductive inferences about novel categories when they contained familiar properties. However, these studies may only reflect how a child reacts within the experimental setting, and thus lack external validity. When met with an object in 'real life' which has a feature correlation which is inconsistent with their current theories, the child may choose to either ignore the contradicting feature correlation, or be motivated to explain it by seeking out information which could lead to the child modifying their own present theory (Barrett et al., 1993).

In the above experiments, the term 'prior knowledge' is used to encompass children's expectations or beliefs about concepts. A question arising from these experiments is the origin of the children's prior beliefs and knowledge of categories (Carmichael & Hayes, 2001). One theory is that children acquire conceptual knowledge through generating 'naïve theories'. What constitutes a 'naïve theory' has been debated, but this term is often used to describe common sense theories which incorporate concepts (Murphy, 1993). Naïve theories provide organised and fairly coherent sets of knowledge about something which enables a child to provide causal descriptions for phenomena and, from these, form logical predictions about unfamiliar objects or events (Keil, Smith, Simons, & Levin, 1998). In other words, concepts are acquired and refined through reference to a child's existing concepts.

However, while much research has focused on how children self-generate concepts, much of their knowledge is also derived from the input of others (Gelman, 2009). This includes learning through teachers and parents, (Callanan & Oakes, 1989).

It is not always possible to pinpoint exactly where knowledge was acquired. McShane and Dockrell (1983) state that it may be "virtually impossible to know in what contexts a child will have acquired a word and therefore it will be impossible to predict patterns of results in advance of a comprehension test..." (p. 65). Despite this, the research in this sub-section demonstrates that both prior knowledge and novel stimuli are central to category acquisition and refinement. These abilities are observed in children from a young age; however, some debate exists around the developmental shift in a child's dependence on these different sources of information. Farrar et al. (1992) propose that learning new concepts becomes more theory-driven as children accumulate knowledge about a particular domain. This idea is similar to Keil's (1989) view of a developmental shift in categorisation from focusing on broad characteristics to defining finer differentiations within these categories. However, these developmental shifts may not be linear; one's learning of concepts may also get constrained by a person's pre-existing knowledge from an early point in their development (Keil, et al., 1998).

1.4.1.3. *When do concept formation abilities manifest?* - research into when concept formation abilities develop is an area of some debate; however, there is much research that attests to young children having concept formation abilities. One conventional approach to gaining a deeper understanding of when these concept formation abilities develop in children is through looking at research on the developmental trajectory in 'typically developing' children (see Table 1).

According to classical empiricism, concept formation begins in infancy as knowledge derives from our senses (Charlesworth & Lind, 2010). When infants explore their environment they gain a sense of shape, weight, and size. Shapes of objects are learned from the experiences of touching. Concepts can therefore be viewed as direct representations of a child's sensory and perceptual experience, or combinations of both (Gelman, 2009). Many researchers find the empiricist approach to learning appealing as it provides the potential to view mental processes and knowledge as being built up from simple building blocks (Gelman, 1996). Similarly, Smith et al. (1996) proposed that one's higher-level cognitions result from lower-level associations; building up sensory cues gradually and throughout many lived examples. Mandler (2008) studied how our earliest, unstructured concepts are sensory and formed as infants. This fits with the sensorimotor stage within Piaget's Theory of Cognitive Development (1954, 1964; see Table 1).

Nativist approaches to concept acquisition offer an alternative view. This approach describes how any concept acquisition involves some innate capacity, even within empiricist approaches (Wanner & Gleitman, 1983). However, the questions that arise from this approach are: how rich and detailed concept formation based on our innate capacities can be, the degree to which someone can incorporate variable concepts from different contexts, and how much these concepts are then modified by their environment (Murphy, 1993). Nativist theories range from children having different attentional biases due to their innate perceptual limits (Mandler, 2004) to a more extreme claim that all concepts are innate (Fodor, 1981). However, whether knowledge is initially derived through our

Table 1.

*Developmental Trajectory of Concept Formation*

	<2 years	2-4 years	4-8 years	8-14 years
<b>Concept Formation Abilities</b>	<p><b>Six months -</b> distinguishes between types of objects. <b>Nine months -</b> can group objects (Starkey, 1981).</p>	<p><b>Two years -</b> children readily make category distinctions, although early categorisations are formed without planning and self-monitoring which are characteristic of older children and adults (Starkey, 1981).</p>	<p><b>Four - 5 years -</b> significant progress in flexibility of mental set; able to observe two sorting criteria in same group of objects (Jacques &amp; Zelazo, 2001).</p>	<p>Child is in 'concrete operational' stage (<b>7-12 years</b>; Piaget, 1954, 1964), characterised by logical thinking about concrete things.</p>
	<p><b>One year and 6 months - 2 years -</b> can categorise objects on basis of function (Nelson, 1973). These children did not demonstrate knowledge of category labels, which contrasts with the traditional view that language is a necessary precursor of concept development (Brown, 1956).</p>	<p>Clustering is a basic and automatic activity in verbal organisation at this age (Rossi &amp; Rossi, 1965).</p> <p><b>Two and 6 months - 3 years -</b> Goldberg, Perlmutter, and Myers (1974) found after presenting short two-word lists to children, related items were recalled more readily than unrelated items, suggesting children make use of verbal categories at the onset of language (Starkey, 1981).</p>	<p><b>Five - 7 years -</b> ability to define a new (third) classification criterion (Jacques &amp; Zelazo, 2001; Smidts, Jacobs, &amp; Anderson, 2004). However, difficulty identifying a second / third criterion without a clearly defined task structure (Smidts et al., 2004).</p>	<p><b>Eight or 9 -</b> an integral part of a child's cognitive potential is determining common factors in sets of different objects enabling them to sort or group objects into classes and subclasses (Inhelder &amp; Piaget, 1964). Abilities in forming and flexibly changing concepts are very important in developing academic skills (van der Sluis, de Jong, &amp; van der Leij, 2007). This level of concept formation peaks aged <b>8 to 9</b>. Afterwards it is gradually replaced by other, more mature, concept classifications, which appear to reach relative maturity at around 11 years (Reichard et al., 1943).</p>
	<p><b>One year, 18 and 24 months -</b> spontaneously manipulate objects in a way that demonstrates ordering abilities (Ricciuti, 1965).</p>	<p>Preoperational stage begins (<b>2-7 years</b>; Piaget 1954, 1964) children begin to learn new rules through labelling objects and make believe play.</p>	<p><b>Five - 8 years -</b> children go from mainly thematic relations early in their development to using taxonomic relations in their school years. By <b>8 -</b> adulthood, taxonomic relations are used much more frequently (Hashimoto, McGregor and Graham, 2007). These relations continue to become less susceptible to task demands, more varied, and more robust.</p>	<p>Concept formation abilities are enhanced: e.g. Barrett et al. (1993) found that <b>9-11 year olds</b> were significantly better than <b>6-10 year olds</b> on a task of category membership (including features that were consistent with or contrasting to their category knowledge). Bekman, Goldman, Worley, and Anderson (2011) found abilities increased between <b>8-12 year olds</b>; they hypothesized enhanced concept formation skills may allow young people to integrate a broader range of potentially inconsistent information, including incorporating newer, positive information with existing negative information, and information from distal sources, (e.g. media, peers), rather than just personal experience (Bekman et al., 2011).</p>
	<p><b>Up to 2 years -</b> sensorimotor stage (Piaget, 1954, 1964) children learn through sensory environment; repeating patterns of movement or sound.</p>	<p><b>Three years -</b> in minimally demanding tasks, children are highly accurate at identifying taxonomic relations (e.g. categorizing a picture of a zebra with either a tiger or balloon; Scott, Greenfield, &amp; Urbano 1985). But when task demands are high and involve more abstract taxonomic relations the high performance declines.</p>	<p><b>Eight years -</b> majority of children can shift between different grouping principles (Reichard, Schneider, &amp; Rapaport, 1943).</p>	<p><b>Twelve-14 years -</b> frontal lobes (associated with concept formation abilities, Morris et al., 1993) are the final neurological structure to mature (Lord-Maes &amp; Obrzut, 1997).</p>

senses (Harris & Koenig, 2006) or innate abilities; Gelman (2002) proposed that it gets elaborated with a person's experiences and development.

As a child develops they tend to structure concepts around either themes or taxonomies. Thematic relationships group together objects that are found together, for example paper and pen may both be part of a classroom theme, as are desk, computer, pupil, and teacher. Lin and Murphy (2001) propose that thematic relationships may be: functional (e.g. pen and paper), causal (e.g. pen and teacher), or spatial (e.g. paper and desk).

Taxonomical concepts can be described by features which share an essence; Rosch (1978) described how basic concepts within our environment can be described through co-occurring features that cluster together ('feature correlations'), such as 'beak' and 'feathers' occur together in a bird. Malt and Smith (1984) stated that using a correlational structure of 'features', or the essence of characteristics, is one way of linking together knowledge about an object which forms the basis of basic level concepts.

Inhelder and Piaget (1964) observed that pre-schoolers arranged categories of objects in accordance to their spatial (thematic) relationship rather than their perceptual similarity (taxonomic). Pre-schoolers also demonstrated a tendency to use thematic relations during tasks involving forced choice matching (Smiley & Brown, 1979) and naming tasks (Scott et al., 1985). However, Smiley and Brown (1979) suggested that while 6-7 year olds may prefer to thematically sort objects (e.g. grouping a spider and a web), they are able to categorise pictures taxonomically (e.g. grouping a spider and an ant).

In contrast to the above theories, some researchers believe that it is not until a child is in their early adolescence that they have sufficient understanding of the relationships between category and subcategory to complete tasks that require concept formation abilities. For example, in order to complete complex sorting tasks, including "Vygotksy blocks" (Piaget, 1964; Vygotksy, 1986). This is because young children are said by these researchers to lack the 'concept' of a group or category (Starkey, 1979). However, this refers to advanced, conceptual

level reasoning to understand higher, taxonomical 'levels' of concepts and related taxonomic principles, including ordinate, superordinate, and subordinate classifications (e.g. a Labrador is a dog, a dog is an animal, and a cat is also an animal). A number of studies have since criticised Piaget's work for underestimating children's abilities in relation to their age. For example, Wood, Smith, and Grossniklaus (2001) stated Piaget used overly complicated tasks, with unclear instructions, and that children could correctly complete simpler forms of such tasks which require the same concept formation and problem solving abilities. Therefore, it would be a mistake to assume that because a young child could not perform advanced or complex cognitive tasks requiring what Piaget (1964) called 'formal operations' that they are not actively engaged in making category distinctions (Starkey, 1979). This suggests that young children do have concept formation abilities, but for these to be tested, tasks must have clear instructions and use a structure which is suited to the developmental level of the child.

*1.4.1.4. Dichotomous divide between adults and children?* - other key questions in designing a test specifically for children are whether and how children's and adults' concept formation abilities differ. In several traditional explanations, concept formation is said to undertake qualitative, fundamental shifts throughout development.

Children and adults are frequently viewed as occupying opposite ends of several developmental dichotomies, including: concrete to abstract (Piaget, 1951), perceptual to conceptual (Bruner, Olver, & Greenfield, 1966), or similarity to theories (Quine, 1977). Dichotomies such as these are appealing, particularly as children can appear to reason in strikingly dissimilar ways to adults. For example, children's and adults' concepts may differ in the number or salience of components they integrate. Such differences may result from the amount of domain-specific knowledge each possesses. These ideas have been supported by research which seemingly demonstrates children often being "prone to accept things as they seem to be, in terms of their outer, perceptual, phenomenal, on-the-surface characteristics" (Flavell, 1977, p. 79). For example, in Piaget's 'conservation error' (Piaget & Szeminska, 1952), where liquid was transferred

from a wide glass into a tall, narrow glass, children below six or seven years old appeared to mainly attend to one salient but deceptive element (e.g. height) and perceive that the transformation led to an increase in quantity. However, there are many critiques of this task, such as Rose and Blank (1974) who found many six year olds answered correctly in their version of this study. Additionally, the children may have assumed that the adult deliberately changing the appearance of the water was important and thus give an answer which was impacted by experimenter demand. Therefore, McGarrigle and Donaldson (1974) devised a task where the alteration was 'accidental' in their 'naughty teddy tasks', which involved two identical rows of sweets being lined up, and a naughty teddy 'accidentally messing up' one line. In these tasks, more than half of the children aged 4-6 years were able to identify that there were still the same number of items in each group, demonstrating that children are able to conserve at an earlier stage than Piaget claimed. This again demonstrates the need for careful design of any test of concept formation for young children.

Various studies argue that developmental dichotomies such as the shift from 'perceptual-to-conceptual' give an incomplete view of children's capabilities (Gelman, 1978; Markman & Hutchinson, 1984; Nelson, 1977). Furthermore, traditional assumptions such as viewing children's concepts as uniformly concrete and perceptual while viewing adult's concepts as conceptual, logical, and abstract (Gelman & Au, 1996) do not adequately capture a child's concept formation abilities. For example, while Inhelder and Piaget's (1964) hierarchical classifications (the ability to simultaneously classify objects into specific and more general groups) required a child to be at the 'concrete operational' stage (7-11 years), other studies have found that using hierarchical structures in categorisation can be detected in the first few years of life (e.g. Nelson, 1973). The development of more sensitive versions of Piaget's tests led to children being able to succeed on such tasks, and thus led to these assumptions about children's concept formation abilities being questioned (Gelman & Baillargeon, 1983). Therefore children are also capable of arranging categories into hierarchical systems that include both abstract and specific categories (Gelman & Au, 1996). Thus rather than just being domain specific and context dependent (Gelman & Baillargeon, 1983), typically developing children's abilities to form

concepts enables them to think beyond exterior features in their categorisations and can also help to decide which components can be applied to other members of that category.

In summary, there is a complex relationship between a child's developing concepts and the way that adults develop and use concepts. While children can form categories and concepts in some similar ways to adults, these abilities develop over time. The age at which a child is viewed to develop concept formation abilities may depend largely on which of the many methodological approaches is taken. Part of the problem appears to rest with whether the methodology of the tests is appropriately accessible to, and developed for, the age of the examinee. However, what is clear is that many studies demonstrate that children have concept formation abilities from as young as infancy which develop and change with age (e.g. Albert, Opwis, Regard, 2010).

As the ability to generate and flexibly change concepts are crucial abilities in academic and adaptive skills (Gligorović & Buha, 2013), it is important to develop an appropriately sensitive test aimed at a child's developmental level which is able to identify whether there are constraints or deficits in a child's concept formation abilities. Deficits in concept formation abilities in simple contexts are predictive of how someone plans and approaches problems in other, more complex situations (Gligorović & Buha, 2013). Therefore, by identifying difficulties early, this means that clinicians and teachers can implement appropriate interventions to help towards a child's ability in effectively navigating the world in day-to-day life, education, and future work (Gligorović & Buha, 2013).

In discussing studies which assess concept formation abilities in children thus far, it is clear that children have concept formation abilities from a very young age which some tests may not be sensitive enough to usefully assess. The next subsection explores the more widely used, current tests of concept formation in more detail.

#### 1.4.2. Assessment of Concept Formation

In order to situate this present study within the broader subject matter, an overview of current and previous neuropsychological assessment methods of concept formation is given in this sub-section.

One key distinction within means of assessing concept formation is between tests that use a single trial versus a multi-trial format. In the single trial format, everything you need to solve the task and derive the concept is provided in the trial. Multi-trial formats in contrast require participants to develop answers through multiple trials within the same task.

*1.4.2.1. Single trial tests* - single trial tests may be either visual or verbal in nature.

*1.4.2.1.1. Visual single trial tests* - the classic example is the Cattell Culture-Fair Intelligence Test (Cattell & Cattell, 1973) which was originally conceived in the 1920s (Sweetland & Keyser, 1991) but is still one of the most frequently used tests of non-verbal intelligence (Urbina, 2011). This test is used for children aged four through to adults, and was designed to be an unbiased, culturally fair test of reasoning using pictures rather than words. It consists of four reasoning tasks, including 'matrices'. The most popular test of this kind is currently Raven's Progressive Matrices (Raven, 1936; Raven, Raven, & Court, 1998), where a person is shown a series of patterns and is required to identify the missing pattern. There have been many parallels of this task, including those used in the 11+ exams (for grammar school entry) and in the Civil Service Entrance Exam. Versions of this are also used in tests such as the Leiter International Performance Scale, Third Edition (Leiter-3; Roid, Miller, Pomplun, & Koch, 2013) and the Matrix Reasoning task in the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V, Wechsler, 2014) and Wechsler Adult Intelligence Scale - Fourth UK Edition (WAIS-IV, Wechsler, 2008). However, some studies have questioned the extent to which these tests are culturally unbiased, for example, Nenty (1986) administered the Cattell Culture-Fair Intelligence Test to American, Indian, and Nigerian adolescents, and found 59% of the items to be culturally biased in favour of the American participants.

1.4.2.1.2. *Verbal single trial tests* - the Similarities task (in the WISC and WAIS batteries) is classically used as a test of concept formation (Rapaport, 1946). This task requires the examinee to identify the qualitative relationship between two objects or ideas. Other examples include the Analogies task (e.g. Wide Range Intelligence Test; Glutting, Adam, & Sheslow, 2000) or Proverbs task, which was originally by Gorham (1956) with newer versions including the Delis-Kaplan Executive Function System™ version (D-KEFS; Delis, Kaplan, & Kramer, 2001). In the Analogies task, a participant is read an analogy that they have to complete (e.g. “wheels are round and boxes are...”). The Proverbs task requires a participant to describe the meaning of a proverb. When using the Proverbs task, the examiner should keep in mind that familiar proverbs feature more in older generation’s conversations than in younger, and thus at least a few proverbs that are not commonly used should be added (Albert & Knoefel, 2011; Van Lancker, 1990). However, the Proverbs and Analogies tasks are not widely used due to being confounded with vocabulary or general knowledge as it is difficult to score highly without a person having a good grasp of vocabulary.

1.4.2.2. *Multi-trial tests* - there are widely available multi-trial tests of concept formation within test batteries such as D-KEFS and Woodcock-Johnson (WJ-III-Cog; Woodcock, McGrew, & Mather, 2001). These include visual and verbal trials.

1.4.2.2.1. *Visual multi-trial tests* - one type of visual multi-trial task are sorting tasks. The D-KEFS Sorting Task, modelled after the California Card Sorting Task (Greve, Farrell, Besson, & Crouch, 1995), involves *free sort* and *recognition* conditions. The free sort condition involves participants sorting six cards into two groups based on similar visual-spatial or verbal-semantic characteristics as many times as possible and to describe their sorting criteria. The recognition condition involves the examiner sorting the cards and asking the examinee to explain the sorting criteria (Delis, Kramer, Kaplan, & Holdnack, 2004).

Other sorting tasks include the Weigl Color-Form Sorting Tests which was created by Weigl (1941) and incorporated into the Goldstein-Scheerer battery of test (Goldstein & Scheerer, 1941) which can be used across the life span from

infants to elderly people. The Color-Form Sorting test consists of 12 cardboard pieces of four colours and three shapes. The participant is first asked to group those that belong together. After their first sorting group, the participant is asked to sort them a different way (Reichard et al., 1943). Participants are then required to explain their groupings. Success on this task is measured through a participant's ability to shift between different categories; failure is measured by a repetition of the first category, or by mixed or incorrect groupings. This test is not used frequently and little research has been carried out using this test, which is perhaps due to this test not being scaled and thus it does not allow for quantitative comparison or evaluation of test performance (Weiss, 1964). However, it does yield rich observational data on a participant's performance during the task.

The Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993) is a commonly used neuropsychological test of concept formation (Maddox, Filoteo, Glass, & Markman, 2010) for people aged 7-89. Participants are asked to match cards in line with one of three classifying criteria (colour, shape, number of features), but are not told how to match them and the sorting rule changes without warning after 10 successive correct pairings. Participants have to reach a conclusion based on the examiner's reaction to their previous response. Success on this task mainly depends on: the ability to detect and shift to an abstract sorting principle, conceptualisation, inhibition of irrelevant responses, cognitive flexibility, and working memory (Gligorović & Buha, 2013).

In using the WCST with children, some research has indicated that at 10 years, children's abilities in switching contexts and inhibition of irrelevant responses reaches that typically observed in adults (Huizinga & van der Molen, 2007; Welsh, Pennington, & Groisser, 1991); with both typically developing (Chelune & Baer, 1986) and clinic-referred paediatric children (Chelune & Thompson, 1987). However, the ability to 'maintain a set', for which working memory is largely responsible, appears to continue to develop after the age of 10 (Huizinga & van der Molen, 2007). The WCST test is seen as being good at capturing perseverative tendencies in comparison to some verbal tests of concept formation (Baldo, Delis, Wilkins, & Shimamura, 2004). However, some

researchers have questioned its construct validity. For example, Barceló (2001) reports that some of the WCST scores do not inform on the nature of cognitive impairment, a critique which underlies some other criticisms about the WCST lacking reliability and validity in pinpointing damage within the prefrontal cortex (Bowden et al., 1998). That said, the WCST scores are still good at identifying a person's general executive system abilities including a person's attention (Lezak, 1995).

The Brixton Spatial Anticipation Test (Burgess & Shallice 1997) has similarities with the WCST, as examinees have to work out the rules and rule changes based on the examiner's response to their last selection, except this task uses a grid within which placement of a blue dot changes. However, while the WCST can be used with children as well as adults, the Brixton task has not been normed with children. An additional disadvantage of this task, is that the scoring system used involves 'Sten scores' as a metric (meaning that scores correspond to a 0.5 of a standard deviation). Thus there is potential for meaningful differences between raw scores to be obscured.

The Concept Formation subtest of the WJ-III-Cog is normed for people aged 2-95 years. It involves visual puzzles consisting of squares and circles outside and inside a box. Examinees are required to ask questions in order to identify and then verbalise underlying concepts and rules for being included in the box (e.g. colour). This subtest includes items for *single comparison* (one box with only one sorting criterion), *multiple comparisons with a single rule* (multiple boxes with one sorting criterion), and *multiple comparisons, with multiple rules* (multiple boxes and rules, alongside either an "and" or an "or" relational concept, e.g. "Yellow and two"). A strength of this test is the large sample across a wide age range was used to standardize this measure (Edwards & Oakland, 2006). However, the WJ-III-Cog has been reported to have low construct validity and low subtest reliability for some of its subtests (Campbell, Brown, Cavanagh, Vess, & Segall, 2008).

1.4.2.2.2. *Verbal multi-trial tests* - some verbal tasks involve deriving concepts and meaning. In the D-KEFS Word Context test (Delis et al., 2001) examinees are shown a pseudo-word and are read a series of clues from which they must

guess the meaning of the pseudo-word after each clue. Floyd et al. (2006) demonstrated this subtest had a reliable convergent validity with WJ-III-Cog tests in a large sample of adults and children.

The 20-Questions task originated from a game whereby a person asks yes/no questions in order to guess the name of a famous person (Baldo et al., 2004). In neuropsychological testing, this has been adapted to assess executive functions such as abilities in planning, seeking, and organising stimuli while monitoring feedback (Anderson, 1998). This task uses a paper containing pictures which can be categorised into multiple groups (e.g. plants, vehicles), and participants have to try to find a target picture through asking “yes/no” questions. Mosher and Hornsby (1966) were the first to adapt a 20-Questions task into a neurocognitive test (Remine, Care & Brown, 2008) to investigate the types of questions examinees asked, and how examinees utilised feedback to generate strategies. Mosher and Hornsby classified the questions asked as either ‘constraint seeking’ (i.e. attempting to eliminate half of the items), ‘hypothesis seeking’ (involving “a series of questions, each of which tests a self-sufficient specific hypothesis that bears no necessary relation to what has gone before”, Mosher & Hornsby, 1966, p. 88) or ‘pseudo-constraint-seeking’ strategies (i.e. questions that eliminate only one object at a time, but are phrased similarly to ‘constraint seeking’ questions, e.g. “Does it meow?”). Alongside concept formation, this task tests a person’s working memory as they are faced with an unchanging array of pictures and have to remember which items their questions have eliminated. Also, cognitive flexibility is needed as participants are required to adapt their questioning according to a changing information set (Alderson-Day & McGonigle-Chalmers, 2011). This concept formation task differs from other tests by providing a larger amount of items within a hierarchical structure, which provides a wider assessment of the examinee’s categorisation skills (Baldo et al., 2004).

More recently, the D-KEFS 20-Questions task (Delis et al., 2001) was developed for people aged 8-89 years. Similarly to Mosher and Hornsby’s (1966) test, this task assesses concept formation from examinees’ verbal responses during a task involving pictures of items organised in such a way to provide structured categories for grouping (e.g. with the same number of vehicles, plants, and

animals). Examinees must identify categories in order to subdivide the items and eliminate irrelevant items through asking as few yes/no questions (maximum of 20) as possible to guess the correct answer (e.g. “Is it an animal?”, Siegler, 1977) while keeping track of their previous responses and modifying answers based on examiners’ feedback.

A limitation of such tasks is that they rely on a participant’s language abilities, and require a participant to use their working memory to keep track of their previous questions, and which corresponding items have already been eliminated (Alderson-Day & McGonigle-Chalmers, 2011), thus the scores on this task do not just reflect a person’s concept formation abilities. These critiques have led some researchers to make adaptations to this task. Alderson-Day and McGonigle-Chalmers (2011) proposed a 20-Questions task with novel robot characters and allowed participants to physically eliminate irrelevant items to reduce working memory demands. They developed this novel task to test it against a standard 20-Questions task with both children with ASD and typically developing children to assess whether the adapted test was more suitable for children with ASD. Despite some limitations to this task (see section 1.4.3.2.), this adapted version of the 20-Questions task was developed in such a way to be accessible to children and thus it was one of the tests that inspired the creation of the test created in this study (see section 1.6.2.).

Despite some critiques, 20-Questions tasks are popular among tests of concept formation due to being an enjoyable and engaging task. Baldo et al. (2004) found the 20-Questions task is tolerated and well enjoyed by both control participants and those with focal prefrontal lesions, unlike some tests of concept formation which can cause frustration or refusal from the latter group of participants. This task being enjoyable is particularly pertinent to children, as using a task that is more ‘game-like’ can increase engagement (Gioia, 2015).

1.4.2.3. *Considerations* - The standard process in neuropsychological tests of concept formation differs from most other neuropsychological tests. For example, within the 20-Questions tasks, the examiner does not measure whether someone’s question yielded a right or wrong answer; rather these tests focus on

the quality or process of the questions asked (Lezak et al., 2004). However, this follows an assumption that a person's concept formation abilities can be tested through the examiner's qualitative judgments of whether the examinee's answers were: simple or complex, abstract or concrete, apt or irrelevant (Lezak et al., 2004). Additionally, Albert and Knoefel (2011) report that the process of an examiner scoring participants' responses in tests of concept formation can be problematic, as scores may reflect their own experiences, cultural background, and biases, alongside their choice of test. However, this assessment process appears to be in line with current research on concept formation, as it provides a means for examinees to demonstrate their ability to alter their approach based on feedback and problem solve when faced with a range of options.

1.4.2.4. *Principles of good cognitive tests for children* – most of the tests that are widely used to test children's concept formation abilities were originally developed for use with adults. As creating a test of concept formation specifically for children is a new area of research, there are currently no principles to guide how to create such a test. However, some researchers have identified general principles of how to conduct a good cognitive assessment with children. Yeates and Taylor (2001) propose four principles of child neuropsychological assessment based on models of neuropsychological assessment (including Bernstein, 2000; Taylor & Fletcher, 1990; Yeates & Taylor, 1998): adaptation; brain and behaviour; context; and development.

Adaption describes the child's interactions between themselves and their environments. Failing to adapt can result in a child struggling with school work or socially. Neuropsychological assessments should thus help to explain why a child is failing to adapt in order to facilitate outcomes in future at school, home and later life.

An analysis of the relationship between a child's brain and their behaviour can provide an insight into a child's adaption. However, rather than uncritically using an adult model of brain functioning, children's assessments should take into account factors such as age, cognitive skills, and the nature of their reported impairments (Dennis, 1988). One way of doing this is to take steps towards

making the test more suitable for the child's age in order to enhance engagement. Gioia (2015) found that using a game-like procedure can increase engagement in children, yet to the researcher of the present study's knowledge, there are no widely used game-like tests of concept formation that have been designed specifically for children. Other considerations include the test not being too long in order to not exhaust a child's attention span, and making the test fit for the child's size (e.g. making children's blocks not too big).

During an assessment, the examiner must also explore the child's context to rule out any other potential causes for adaptation difficulties. Thus while measuring a child's cognitive skills; the examiner must also determine how children apply these to meet the demands of their own environments in order to make more specific and informed recommendations (Bernstein, 2000).

Behavioural and cognitive development results from an interplay between biology and a child's environment, and involves new skills emerging, such as language, and we lose earlier abilities such as more primitive reflexes (Dennis, 1998). Thus the tests must be appropriate to a child's developmental level. The test instructions should be written using accessible language that the children understand, and they should be suited to the developmental level of the child (e.g. McGarrigle and Donaldson, 1974).

#### 1.4.3. Aetiology of Difficulties

There are several potential causes of executive dysfunction in children which can be separated into two main sources: brain damage (Elliott, 2003) and congenital disorders (Friedman et al., 2008). Below is a discussion of brain injury and pervasive developmental disorders which are two areas which cover some of the more common reasons for deficits in concept formation in children.

1.4.3.1. *Brain injury* - the frontal lobes, specifically the lateral and dorsal parts of the prefrontal cortex, have been specifically associated with problem solving and concept formation (Morris, et al., 1993). The frontal lobes can be damaged by any form of acquired brain injury, including a stroke, tumour, and meningitis

(Headway, 2017). They are also especially vulnerable to Traumatic Brain Injury (TBI), due to their large size and location at the front of the brain. Even a knock to the back of the head can lead to frontal lobe injury due to the brain moving within the skull as a result of the impact (Headway, 2017). Fairly common sequelae in children with TBI includes inattention and impairments in problem solving (Donders & Kuldanek, 1998). Other researchers have indicated that conceptual functions tend to be impacted after a brain injury regardless of the site of the injury (Luria, 1966; Mesulam, 2000).

These deficits have been demonstrated in research using neurocognitive testing. Burgess and Shallice (1997) found a highly significant difference between the performance of adults with frontal lesions and controls on the Brixton Spatial Anticipation Test. Baldo et al. (2004) found in their study using the D-KEFS 20-Questions task, that people with pre-frontal lesions asked significantly more questions than the control group and sometimes exceeded the 20 questions limit. These participants tended to use less effective category selection (e.g. asking a question relating to only one item), and were more likely to use repetitive questions (3%) compared to controls (1%). This study only used a relatively small number of brain injured participants (n=12) due to a strict inclusion criterion (including normal language abilities, focal frontal lesion). This was deemed necessary to provide consistency within their population to test for difficulties in concept formation in the absence of other factors such as language dysfunction or memory deficits.

Kizilbash and Donders (1999) tested the performance of 9-16 year old children who had a TBI diagnosis on the WCST. They found an association between younger age and longer duration of coma with poorer performance. However, not every child with a TBI performed badly on the WCST; a third of the children completed the test.

1.4.3.2. *Pervasive developmental disorders* - a range of pervasive developmental disorders have been associated with deficits in concept formation. The most prominently studied example is Autistic Spectrum Disorder (ASD). A distinctive characteristic of ASD is having intense preoccupation with specific information

categories (Frith, 2003) such as train timetables. These interests traditionally fall within the 'restricted interests and repetitive behaviours' third of the autistic 'triad of impairments' (Wing, 1996). The categorisation process is fundamental in efficient information processing (Rosch, 1978), and therefore difficulties in dealing with departure from familiarity or fixations on particular categories could perhaps reflect abnormal or deficient categorisation processes. Many studies document that individuals with ASD have difficulties with tasks testing concept formation, such as reduced completion of category sorting on the WCST (Lopez, Lincoln, Ozonoff, & Lai, 2005; Shu, Lung, Tien, & Chen, 2001). It has been suggested that the typically observed restricted and repetitive characteristics in people with ASD could be due to difficulties with category generation, or biases in forming and using categories (Klinger & Dawson 2001). Ropar and Peebles (2007) found individuals with ASD had sorting preferences for concrete over abstract categories. Much research has indicated that people with ASD complete significantly less trials than controls in the 20-Questions task and tended to use questions that eliminated single items rather than identifying and eliminating categories (Minshew, Siegel, Goldstein, & Weldy, 1994).

These findings suggest that there may be a fundamental problem in concept formation in people with ASD (Minshew, Meyer, & Goldstein, 2002). Although people with ASD seem to recognise and learn category groupings similarly to those without ASD, they may have difficulties generating categories as an information processing strategy (Alderson-Day & McGonigle-Chalmers, 2011). Interpreting reasons for impaired performance on 20-Questions tasks can, however, be problematic as these tests pose various complex demands, such as use of working memory. As working memory deficits have been widely reported in people with ASD (e.g. Williams, Goldstein, & Minshew, 2006), this could contribute to problems on this task.

Alderson-Day and McGonigle-Chalmers (2011) compared results from a standard 20-Questions task (based on Mosher & Hornsby's 1966 and D-KEFS 20-Questions tasks) against a novel, adapted 20-Questions task of artificial items which allowed physical elimination of items to minimise the working memory demands within trials. They found the ASD participants asked significantly less

questions during the novel task, when they could physically eliminate items, than the standard 20-Questions task. Overall participants with ASD asked 'lower quality' questions than controls on both tasks (i.e. using category groupings which eliminated fewer items). They also found ASD participants performed worse than controls in completing a standard version of the 20-Questions task. They were unable to assess overall performance on their novel task due to ceiling effects. Further limitations include a small sample size (n=14 in each group) and the children in each group were not matched on their verbal ability. These limitations mean it is unclear the extent to which conclusions can be generalised from this study (Alderson-Day & McGonigle-Chalmers, 2011). However, this study still provided useful information regarding how people with ASD problem solve and use concepts compared with controls. The findings also indicate that allowing physical elimination of items appeared to usefully reduce the working memory demands on the participants.

#### 1.4.4. Impact of Deficits

The impact of having concept formation deficits is multifactorial, from impacting a child at school, to their psycho-social development. This next sub-section will explore these impacts in more detail.

*1.4.4.1. Academic difficulties* – concept formation is necessary for the development of various academic and adaptive skills (Gligorović & Buha, 2013) including reading, writing, and mathematics. Some research studies have indicated a strong association between children's conceptual skills and their performance on tests of vocabulary knowledge (Langer, 1967) and reading comprehension (Braun, 1963). While reading, conceptual thought underlies formation of complex relationships between characters and understanding of the story's plot (Guthrie et al., 2004), alongside giving an overall gestalt of the purposes for the reading material. Expressing ideas through writing also involves conceptual organisation of ideas, in order to write in a coherent way while adhering to a principal theme.

Concept formation and problem solving abilities are also used in a range of mathematical processes including operations, such as subtraction, and in understanding what these operations and the applications of these represent.

Research in this area has indicated that a child's conceptual mathematical skills when they start school are highly predictive of their later academic success (Duncan et al., 2007). Additionally, both a child's literacy and mathematical conceptual abilities in early life have been shown to be correlated with long-term influences on their quality of life, such as future employment and salaries (Bynner, 1997; Rivera-Batiz, 1992). In fact, concept formation, i.e. the realisation of what objects belong together, seems to be a key part of every intellectual function (Reichard et al., 1943). Therefore, it is not surprising that many intelligence tests include testing a person's concept formation abilities (Wechsler et al., 2004).

The importance of concept formation in academic abilities was demonstrated in a case study by Levi (1965). Levi described the yearlong treatment of an 11-year-old boy, Stephen, who was referred for reasons including long-term 'poor schoolwork'. Stephen's intervention focused on developing his concept formation abilities through three phases: learning to generalise, learning about relations between concepts, and being able to notice conceptual relations. Afterwards, Stephen's WISC scores increased, most notably his 'Similarities' score (from a scaled score of 7 to 14), which is a test considered to be a measure of concept formation (Rapaport, 1946). Stephen's school grades also increased to being within the average range for his age, and a 6 months follow-up showed his school performance had continued at this level. It is unclear how generalizable the results from this study are as this used a case study of just one child, and perhaps improvements were made in part due to Stephen having increased attention. However, developing teaching methods for children with difficulties in perception and concept formation is not a new idea; support for this study comes from Albert A. Strauss who helped rehabilitate children with such difficulties who had previously found it difficult to learn in conventional school classrooms to then be able to return to these school settings (Gardiner, 1958).

From their study with children with mild intellectual difficulties (MID), Gligorović and Buha (2013) concluded that poor WCST results may result from children with MID having difficulties in generating a new pattern of concepts, task conceptualisation or shifting between mental sets. They concluded that progress in conceptual abilities is achieved when a child develops their abilities in mental set flexibility. Thus, they recommended that interventions should offer activities that are specifically designed to vary in their actions, attentional demand, conceptual patterns, and content. These findings can only be generalised to children aged 10-13 years old, as this was the sample used in this study. However, this study provided an important suggestion that in order to test concept formation abilities, a test should be able to capture a child's ability to flexibly switch between sets, such as a task which requires children to adapt their answers based on examiner feedback.

*1.4.4.2. Psycho-social difficulties* - a child's concept formation abilities are also important for information processing outside of the academic domain. Conceptual thinking is crucial in social interactions in order to comprehend and relate to others' points of view (Castellanos et al., 2015). Verbal communication relies upon concepts being transmitted through language. Thus, concept formation abilities play an integral role in a child developing adequate verbal communication skills. Whether in a playground, or presenting to classmates in a class room, these skills are guided by general concepts which involve the organisation of materials, ideas, and planning, which are key foundational components within executive functioning (Gioia, Isquith, Guy, & Kenworthy, 2000).

Gioia et al. (2000) posit that concept formation is a core foundational cognitive function in everyday functioning which underlies abilities in forming, applying, and manipulating concepts. Due to concept formation being an integral part of adaptive functioning in various domains, Castellanos et al. (2015) described in their study of prelingually deaf children with cochlear implants how identifying, understanding, and reinforcing the development of concept formation abilities among vulnerable groups of children is of crucial importance. Their study found that these adult's and children's performances were significantly worse on the

WJ-III-Cog test than normally hearing participants. However, as a cross-sectional, correlational design was used and relatively few deaf children with cochlear implants were recruited, it is not clear the extent to which the findings can be generalised.

This sub-section has described the importance of concept formation abilities in a child's everyday life. It has also discussed some research demonstrating that concept formation difficulties can be helped and improved. Thus if we were able to identify and provide appropriate interventions early this could help children improve their day-to-day functioning and long term quality of life.

### **1.5. Summary of Literature Review**

From the literature discussed in the above sub-section, it can be seen that the main, current tests of concept formation used with children (usually aged 7 years and above) were developed for adults. However, crucial developments of concept formation abilities begin very early in a child's life (see section 1.4.1.3.). Additionally, neurological development is thought to continue throughout childhood, with the frontal lobes (associated with concept formation abilities, Morris et al., 1993) being the final neurological structure to mature at approximately 12-14 years (Lord-Maes & Obrzut, 1997). Therefore, it is important to also be able to capture concept formation abilities as they are developing. Some research has suggested that children and adults reason in different ways, with children's concept formation abilities being underdeveloped in children (e.g. Piaget, 1964; Starkey, 1979). However, this has been questioned by the development of more sensitive tasks (Gelman & Baillargeon, 1983), including those using accessible language, instructions suited to the developmental level of the child, and a more sensitive research design (e.g. McGarrigle and Donaldson, 1974) in order to more accurately assess a child's concept formation abilities (Wood et al., 2001).

Another reason why it is important to develop a test specifically for children is because adults and children may engage more with different types of tests. For

example; with adults, the face validity of the task (Nevo, 1985) is important for engagement (i.e. the test should look valid, as well as being valid). However, for children, using a game-like procedure can increase engagement (Gioia, 2015). Channon and Crawford stated in 1999 that there were very few cognitive tests that had been designed specifically for concept formation in children; having reviewed the literature, this is still the case. While some tests of concept formation have been adapted for children, there are no game-like tests of concept formation being widely used, which is a weakness given that children are reportedly more engaged more by this type of task (Gioia, 2015). Therefore, based on previous literature, in order to create a test of children's concept formation abilities in a way that will give the best indication of a child's potential, this study must create a test that is specifically designed for children by using a game-like task with instructions that are easily accessible to the child's developmental level.

## **1.6. Present Study**

### **1.6.1. Aims and Rationale**

This current, exploratory study is the first phase in the development of a new neuropsychological test of concept formation; *the Alien Game*. The Alien Game was created with an aim of being an enjoyable and engaging psychometric tool for children, with clear instructions, through using a game-like procedure. A benefit to this design is that children (particularly those with difficulties in executive functioning) may struggle with the format of testing due to difficulties in maintaining attention (Marton, 2008). This study was conducted on 'normally developing' children, in order to see whether it could be possible to generate 'norms' from this population's responses on this task, and whether children found it more or less engaging than other tests of concept formation (Similarities and Matrix Reasoning WISC-IV subtests). If the results found it was possible to generate norms, and that the children found it engaging, then future research could continue the development of creating a standardised version of this task

which could be used to assess and identify deficits in concept formation in children earlier, providing an opportunity for earlier clinical intervention.

#### 1.6.2. A Game-Like Procedure

Games such as 'Guess Who?' are conceptually similar to the 20-Questions task and WCST in that they involve a participant abstracting a category, keeping track of their questions and responses, and modifying their questions based on feedback (Baldo et al., 2004). In the original version of 'Guess Who?' two participants select a character and ask yes/no questions in turn to try to guess their opponent's character, and therefore win the game.

Ideas for the Alien Game were developed from Mosher and Hornsby (1966) and Alderson-Day and McGonigle-Chalmers (2011) 20-Questions tasks, and 'Guess Who?'. Alderson-Day and McGonigle-Chalmers (2011) created a version of the 20-Questions task that used pictures of 24 robots on a hinged board. This made the task more similar to games such as 'Guess Who?' and thus the researcher felt that this made the task more accessible and enjoyable for children. An original version of 'Guess Who?' was not used for a number of reasons. Firstly, as some children may be familiar with the 'Guess Who?' game, the pictures in this task were changed to custom-drawn 'Aliens' (Appendix A) thus reducing the chance of 'practice effects' (e.g. participants familiar with 'Guess Who?' may have a strategy and be less anxious about the task than someone who has not played 'Guess Who?' before). 'Aliens' were chosen to appeal to younger children. The pictures of Aliens, with made up names, were also chosen to make the test more culturally fair, as the pictures were novel to all participants. Additionally, whereas traditional 20-Questions tasks allow items to be categorised in line with conceptual criteria (e.g. animals, vehicles, tools, etc.), which can be culturally laden, this task only required perceptual categories to be utilised. The perceptual characteristics (e.g. colours and shapes) were organised in such a way to provide structured, category groupings.

The game used a board with hinged frames. This was to firstly make it more 'game-like', but also to enable participants to be able to physically remove items by putting down the eliminated items after questions, to minimise the working

memory demands of this test (Alderson-Day & McGonigle-Chalmers, 2011). Unlike 'Guess Who?', the 'Alien Game' was only played one way (with the participant asking questions), to make it similar to the 20-Questions tasks, while ruling out the possibility that the examinee's questions may be affected by the examiner's questions.

Finally, as the task is highly verbal in nature, there is a chance that a deficit in language abilities could impact upon a person's score (Alderson-Day & McGonigle-Chalmers, 2011). As language demands of this task could not be minimised easily by modifying the current procedure, an examinee's verbal abilities were measured by administering the Similarities task (WISC-IV) to test whether any participants had any verbal deficits.

Similarly, the Matrix Reasoning task was administered to test whether any children had any visuo-spatial difficulties. The Similarities and Matrix Reasoning tasks were selected as they are both tests of concept formation, but as they use a different format to the Alien Game (a single trial format rather than a multi-trial) they would not lead to the child's scores being affected by practice effects.

## **1.7. Research Questions**

The present research study aimed to answer the following questions:

- 1) Do children from a non-clinical sample adopt similar strategies when completing the Alien Game?
- 2) Can we establish normative performance characteristics (including scores and typical patterns of responding) from the results that capture normal variation of concept formation?
- 3) Do children engage better with the Alien Game than they do with other standardised measures of concept formation?

## **2. EPISTEMOLOGY & METHODOLOGY**

As the epistemological position influences all aspects of research including the chosen methods, analysis, and interpretations, it is important that researchers explicitly state their epistemological position and subsequent assumptions (Willig, 2012). In stating the epistemological position, the researcher also allows for a consideration of the potential limitations in adopting this approach. Within this present research, a critical realist epistemological position was adopted. A discussion on the different epistemological positions and why a critical realist approach was chosen are explored below.

### **2.1. Epistemological Approaches**

Epistemology relates to the theory of knowledge (Ferrier, 1854). There are various epistemological stances which can be grouped into: realist, phenomenological, and social constructionist (Willig, 2012). Each of these epistemological positions has a particular set of assumptions about how knowledge is constructed, and the relationship between knowledge and ideas about concepts of fact, truth, belief, and subjectivity (Armstrong, 1973). In taking a position of 'realism', a researcher would aim to establish 'real' knowledge about a world that already exists independently from our awareness of it. There are two branches within adopting a realism approach; one is more direct, whereby knowledge is seen as directly mirroring 'facts'; the other is more critical, whereby the researcher believes there is a reality which is 'real' and thus measurable but also views knowledge as being flawed by external social influences and through the imperfections of the researcher's subjective attempts to understand it. A researcher taking a 'phenomenology' epistemology would aim to understand a person's subjective version of their experiences and how they perceived and interpreted an event, as shaped by the researcher's experiences. Thus, researchers aim to understand an experience through the eyes of their participants, without trying to establish the accuracy of an account or to make attempts to understand processes that underlie a person's experience. Finally,

social constructionism focuses on how a person's version of reality is socially constructed through language. Within this position, language and social interaction are viewed as mediating human experience (Willig, 2013). Therefore, the researcher's focus is on how reality is constructed through developing and establishing social discourses and understanding what impact these discourses have on a person's experience.

## **2.2. Adopting a Critical Realist Approach**

As mentioned above, this present research study, and the research questions, are underpinned by an epistemological and ontological position of critical realism. In doing so, this research proposes a co-existence of both the world being 'real' and a person's subjective account being derived from their own context, within their wider historical, political, and social context (Bhaskar, 1989). Within this study, attempts to explore, assess, and quantify certain phenomena (e.g. 'concept formation') were made within a material and social reality which the researcher believes exists separately and independently to personal experience and across time. In taking a critical realist stance, a theory-driven method can be adopted and it is anticipated that this study's results and interpretations will have utility in neurological assessment of 'concept formation' and 'executive functioning' in children in the future.

Despite the above discussion, the researcher concurrently believes that concepts such as 'concept formation', 'abstraction', and 'executive functioning' are not objectively 'real' entities but rather they are socially constructed classifications. Thus, how 'normal' cognitive performance is perceived tends to vary throughout time due to being influenced by historical, cultural, and socio-political contexts (Flynn, 1987). Even the medical diagnoses discussed within this thesis can be challenged. In line with this belief, this research was conducted from a position of not believing that by attempting to measure 'concept formation' that it will find a 'real' absolute truth. The interpretations made should instead be located within the present historical and social context which produces material realities (Harper, 2011).

While analysing and coding the data, the researcher was aware that observations and interpretations made during the task may encompass biases and human error (Trochim, 2000). Therefore, this research was conducted from a position of not being able to take anything for granted and knowledge was positioned as fallible and therefore cannot be viewed as 'fact'. Thus, results and interpretations from this study should be interpreted cautiously while being mindful of their boundaries and limitations.

### **2.3. Inclusion & Exclusion Criteria**

The inclusion criteria were: children aged 6-11 years who understand verbal and written information in English. The lower age limit was chosen to match norms of the Wechsler Intelligence Scale for Children 4th edition (WISC-IV). Therefore, if this study included children who were younger, the researcher would not have been able to have been administered the WISC-IV subtests with these children which could have made it more difficult to compare the children's verbal and non-verbal abilities. The upper limit of 11 years was chosen to capture children's concept formation abilities before a child's frontal lobes mature at around age 12-14 (Lord-Maes & Obrzut, 1997).

The exclusion criteria were: children who have a learning disability or who are known to have had a brain injury. Within the current study, no children were excluded on this basis.

### **2.4. Design**

This study used a mixed methods approach. This research design allowed an exploratory stance to be taken in assessing children's strategies, enjoyability, and performance, while also developing ways to measure the children's responses in this task. The children's responses were measured through both assessing the types and quality of questions asked by the 'typically developing' children, and through quantifying the data into scores that could also be used to explore a

child's abilities in learning to learn. These scores were also compared with children's scores on the Similarities and Matrix Reasoning tasks (WISC-IV).

## 2.5. Materials

This study used:

- Similarities and Matrix Reasoning sub-tests from WISC-IV including scoring sheets
- A Dictaphone
- Pen and paper
- Information sheets (Appendix B, C)
- Consent forms (Appendix B, D, E, F)
- Standardised instructions sheet (Appendix G)
- Likert/visual analogue scale (Appendix H)
- Coding sheet (Appendix I)
- Coding Sheet Explained (Appendix J)
- Scoring for Code 7 and 8 (Appendix K)
- The Alien Game.

In line with Gioia's (2015) finding that a game-like procedure can increase engagement in children, The Alien Game was developed using a game-like structure as a cognitive tool for this task. In order to make the task suitable to the nature of impairments that children with frontal lobe difficulties may experience (Taylor & Fletcher, 1990), the test consists of four short trials as children with these difficulties may struggle with maintain attention (Marton, 2008). The game was also adapted so that the children were also able to physically eliminate items to reduce the demands on their working memory (Alderson-Day & McGonigle-Chalmers, 2011). In order to make the task suitable to the child's developmental level, instructions were written in language that would be accessible to children.

The Alien Game consisted of 24 pictures of 'Aliens' arranged on a board with hinged frames in four rows of six pictures. The Aliens were given made-up, non-culturally specific names and were custom-drawn in order to eliminate any

potential practice effects which may have occurred if the study had used a potentially familiar task (e.g. 'Guess Who?'). Thus, all stimuli were novel for each participant. The three colours of the aliens (grey, yellow, and blue) were chosen so that children who are colour blind would still be able to differentiate between them.

The researcher initially planned to pilot this study. However, due to time restraints, and as only a small number of children were recruited, it was decided to instead recruit all 18 children into the main study and show the initial alien drawings to a group of adults in lieu of piloting the test on children. Based on the feedback given, the tail was made larger, and positioned behind the alien (instead of coming out from the front) as it was reported that this would minimise the chance of the tail possibly being confused with an arm.

As this task used 24 pictures, which is less than other 20-Questions tasks (i.e. D-KEFS 20-Questions task uses 30 pictures), a child was considered to have passed the task if they guessed the target Alien within 10 questions. The rationale for setting a cut-off point for passing the test before the children were tested was to keep this scoring criteria in line with Alderson-Day and McGonigle-Chalmers' (2011) study, which also used 24 pictures. If it was found that this did not appear to be an appropriate cut off point this can be revisited in future studies.

All pictures started standing on the board and the child flipped down those eliminated by their questions. The number of categories and characteristics were based on a combination of Baldo, Delis, Wilkins, and Shimamura's (2004) test of executive functioning which used three levels of categories, and a similar number of characteristic that were used in 'Guess Who?' games. The three category levels included: large categories which comprised of two options; with 10-14 items each (e.g. tail; no legs). The medium categories had three roughly equal groups of 7-9 items each (e.g. colour and shape). The small categories had 3-4 items each (e.g. 4 legs and elephant's trunk).

## 2.6. Ethics

Full ethical approval was gained from University of East London's Research Ethics Committee (Appendix L).

The study was conducted within three schools that were in or near London where the researcher had personal contacts. The Head Teacher of each school was approached to ask whether they agreed to participate in the study. If the Head Teacher agreed, they were also asked whether they would prefer to have the parents' consent to opt their children into the study, or whether they would prefer the parents to opt their child out if they did not wish their child to take part and for the Head Teacher to act in *Loco Parentis* (Appendix F) for the children who remained in. In two of the three schools consent was obtained from the parents being sent information about the study (Appendix C), and being asked to opt their child into the study if they wished via the parental consent form (Appendix E). The third school also sent out the information sheet, but asked the parents to opt their child out of the study if they did not wish them to take part (Appendix D).

For the children who had parental consent, or where the parents had not opted them out, the researcher checked with their teacher that they were eligible for the study based on the inclusion and exclusion criteria. The researcher then met with each participant individually. When the researcher met with the child, the information on the child consent form (Appendix B) was read and discussed with the child to allow an opportunity for questions and to ensure they understood the task before giving consent. If the child agreed to participate, they were asked to sign the child consent form (Appendix B). Both the children and their parents were fully informed about the study and no deception was used. This study did not involve any hazards or risks to participants. Both the parents and the children were advised of their right to withdraw their own/their child's data at any point during the course of the study by contacting the researcher or the child's teacher. Participants were also informed that they could stop the task at any time. If a child became distressed during the task, they would have been offered a break, or asked if they would like to reschedule or withdraw from the study. No participants became distressed during any of the tasks.

### 2.6.1. Confidentiality

Consent forms included the child's name, date of birth, and gender. Each child was then allocated a code which was written on both the bottom of their consent form, and on their observation and WISC-IV scoring sheets. The consent forms were kept separately from the rest of the information (i.e. child's answers to the Alien Game, Similarities or Matrix Reasoning tasks) in a locked draw in the researcher's supervisor's office (which is locked when unattended). This information was kept for the duration of the study in case a parent or child wished to withdraw their consent at any time during the study. After the results had been analysed, the consent forms underwent secure disposal by being shredded in accordance with UEL guidelines. The children's national curriculum data, age, and gender were documented alongside the participant's code and the test data on excel and SPSS databases. These were stored on a password protected laptop. No personally identifiable data (i.e. names) were stored on the computer database.

## **2.7. Procedure**

The test procedure took place in a quiet room in each of the schools. Participants were seated next to the examiner at a desk, approximately 20cm away from the Alien Game. The children were asked whether they had played a 20-Questions game or a version of 'Guess Who?' before. Each child was then read standardised instructions (Appendix G) and asked if they had any questions before the Alien Game began. The child was also advised that they were not to ask questions about the names of the characters. Once the task began, if the child asked a question related to an Alien's name, the participant was informed that this question was not permitted and was reminded, once, that they were to use the fewest possible questions to reach the target answer. As the children flipped down the irrelevant items after each question, the examiner checked that the child had eliminated them all correctly before asking for the next question. Where need be, the researcher prompted the child with a question such as "Is that all of the yellow ones?" The participant's questions were recorded on a

Dictaphone and the researcher's observations were recorded on paper. The Alien Game was run over four trials for each child. The ordering of the same four target Aliens was kept the same for each child to keep the trials consistent.

Participants were then administered the Similarities and Matrix Reasoning sub-tests from the WISC-IV using the Satz-Mogel method (Satz & Mogel, 1962). The Similarities task was administered to screen for whether any child had difficulties in their verbal ability. This was because while the Alien Game was mainly visual in nature, it also involved a verbal component as children had to verbalise their questions to the researcher.

The Matrix Reasoning task was administered to test whether the children had any visuo-spatial difficulties. Both of these sub-tests are single trial tests of concept formation, which present all of the information needed to complete the task to the participant; whereas in the Alien Game the answer is found through asking multiple questions. The WISC-IV tasks depend on abstraction, and therefore reasoning, whereas the Alien Game task requires both of these and effective use of working memory (in order to keep track of what questions were asked before). Therefore, the two WISC-IV subtests were also chosen as while they also assess problem solving abilities, they use a different structure to the Alien Game so the results should not be affected by practice effects.

Participants were asked to complete a five-point Likert scale (Likert, 1932, cited in Carifio & Perla, 2007, pp.106-116; Appendix H) which also incorporated a visual analogue scale for younger children after each task (Alien Game, WISC-IV Similarities and Matrix Reasoning) to rate their enjoyment of the task (Appendix H).

## **2.8. Data Collection and Analysis**

### **2.8.1. Content Analysis**

The recorded data were transcribed orthographically by the researcher. This served as the primary source of data for the two-phase content analysis. A 'coding unit' was defined as each individual question asked by the children.

A two-phase content analysis was carried out on these transcripts. The codes were not concerned with whether the question the child asked yielded a right or wrong answer; they instead related to the quality of the questions asked and whether these questions related to the preceding question. The first phase of generating the coding criteria involved deciding what types of questions the researcher was looking for in advance. This was derived from categories used by previous tests of concept formation (see Table 2). The second phase was data driven and began after the first few participants played the Alien Game. This phase also involved changing or removing some codes from the first phase where necessary after all of the participants had been tested. This was because the Alien Game differs from traditional 20-Questions tasks and therefore required different criteria to assess the types of questions asked. For example, in this task, a child was not allowed to ask about a character's name and the fewest number of items that could be removed at the start was four items, rather than one. Also, this task only required the child to use perceptual information, compared with most 20-Questions tasks which used a mix of conceptual items (e.g. living and inanimate objects). The number of characteristics was also grouped into uneven groupings (e.g. colours were grouped into groups of seven, eight, and nine, rather than three groups of eight) which is similar to 'Guess Who?' order to create a 'game-like' structure with an element of chance.

Once a final coding scheme had been established, the researcher coded the data. Inter-coder reliability was tested by a second rater who was a first time coder, who coded 10% of the data collected using the Coding Sheets (Appendix I). The second rater was also given a Coding Scheme Explained sheet which included: code numbers, category names, detailed definitions of the categories, and examples in order to make the coding categories clear to the coders (Appendix J) and the scoring sheet for codes 7 and 8 (Appendix K). These can also be used to aid the replicability and transferability of the study and the Alien Game to other populations. Much of the coding scheme used in this study has also been extensively recognised in prior research in this area which aids the transferability of the codes used.

Table 2.

Codes From Previous Tests of Concept Formation

	Mosher and Hornsby (1966) 20-Questions	Alderson-Day and McGonigle-Chalmers (2011) 20-Questions Codes followed Minshew et al. (1994).	D-KEFS 20-Questions	WCST	
Codes	<p><b>'Constraint-seeking'</b> - attempting to eliminate half of items by grouping large numbers of possibilities into two domains (e.g. is it living?).</p> <p><b>'Hypothesis-seeking'</b> - questions test "a self-sufficient specific hypothesis that bears no necessary relation to what has gone before" (Mosher &amp; Hornsby, 1966, p. 88), requiring a larger number of questions without depending upon a conceptual scheme.</p> <p><b>'Pseudo-constraint-seeking' strategy</b> - eliminates one item but are phrased like constraint seeking (e.g. "Does it meow?").</p>	<p><b>'Constraint-seeking' questions</b> - referring to two or more items.</p> <p><b>'Hypothesis-scanning' questions</b> - questions that name an item.</p> <p><b>'Pseudo-constraint' questions</b> - referring to one item without naming it.</p> <p><b>'Redundant' questions</b> - eliminating no items.</p> <p><b>'Repeated' questions</b> - the number of repetitive questions.</p> <p>Questions were also coded as either: <b>conceptual</b> (abstract, taxonomic grouping or abstract label; <b>functional</b>, regarding the object's use; or <b>feature-based</b>, regarding its semantic property) or <b>perceptual</b> (e.g. shape).</p>	<p><b>'Spatial questions'</b> - regarding an item's location on the page, e.g. "Is it in the bottom right of the page?").</p> <p><b>'Repeated questions'</b> - the amount of repetitive questions.</p> <p><b>'Set-loss questions'</b> - failure to maintain a cognitive set or to comprehend task instructions.</p> <p><i>Note:</i> These codes are optional measures.</p>	<p>Participants identify sorting categories (<b>'colour'</b>, <b>'form'</b>, and <b>'number'</b>; the code <b>'other'</b> is given to responses that do not match any of the previous three). A participant must determine and maintain a sorting principle or <b>'set'</b>. <b>'Failure to maintain a set'</b> is when an examinee gives five or more consecutive correct answers but then fails to maintain the set until achieving a <b>'category'</b> (i.e. 10 consecutive correct matches). After a category is achieved, the coding strategy changes and the child must resist <b>perseverating</b> with the old principle to determine a new, correct sorting principle. When an examinee persists with an incorrect characteristic, the response is scored as <b>'perseverative'</b>. Answers that were incorrect, but do not match the preservative principle are coded as <b>'nonperseverative'</b>. Responses that match the sorting principle are <b>'correct'</b>; incorrect responses are <b>'errors'</b>. When a response card matches a stimulus card on only one characteristic, the match is <b>'unambiguous'</b>. If the response card matches with more than one stimulus dimension, it is <b>'ambiguous'</b>.</p>	
	Other measures	<p><b>Mean number of questions.</b></p> <p><b>Percent constraint questions.</b></p>	<p><b>Overall percentage of trials completed.</b></p> <p><b>Mean number of questions taken on each trial.</b></p>	<p><b>The total questions</b> - global achievement measure.</p> <p><b>Total weighted achievement</b> - a global measure which accounts for asking a greater or fewer number of questions than the optimal range.</p> <p><b>Initial abstraction scores</b> - minimum number of items eliminated from a person's first question, across four trials.</p>	<p><b>Number of categories completed</b> - number of categories (i.e. a consecutive sequence of 10 correct matches) the client completes.</p> <p><b>Trials to complete first category</b> - number of trials taken to finish the first category, before a person is required to shift the set.</p> <p><b>Percent perseverative errors</b> - number of perseverative errors in relation to test performance overall.</p> <p><b>Percent conceptual level</b> - three consecutive correct responses.</p> <p><b>Learning to learn</b> - average change in efficiency across trials.</p> <p><b>Number of trials administered, total number correct and total number of errors</b> - were also scored for.</p>

### 2.8.2. Analysis of the Coding Data

The content analysis used both a qualitative and quantitative methodology; an approach that Weber (1990) suggested was the best approach to conducting a content analysis. The codes were derived qualitatively through both previous research and from the data itself. Data saturation was achieved as all new data fit into the codes defined and no new codes were needed. The quantitative aspect of the content analysis involved conducting descriptive statistics on how frequently a child's question matched each of the codes. An exploratory content analysis explored the types of questions used and the patterns adopted by the children.

The inter-coder reliability was measured using a Cohen's Kappa Coefficient (Cohen, 1960).

### 2.8.3. Additional Tests

As this is a new test of concept formation, a number of different quantitative measures were also devised and applied in order to establish which were useful measures in capturing and assessing children's concept formation abilities.

2.8.3.1. *Abstraction score* - an 'Abstraction Score' (AS) was calculated for each participant's first three questions across the four trials. The AS was calculated by adding up the minimum possible number of items that could have been removed by these questions regardless of whether or not the question yielded a correct answer. For example, if the child's first question was "Is the Alien blue?", the AS score would be 9, regardless of whether the answer yielded a correct (n=9) or incorrect (n=15) answer. The AS was developed to see whether these children had similar abstraction rates, and was based on the AS used in Baldo et al.'s (2004) test of executive functioning. Only the first three questions of each trial were used in the analysis as some children completed at least one of the trials within three questions.

2.8.3.2. *Learning slope* - in order to assess a child's ability to learn, a Learning Slope for each child was calculated through comparing the AS from Trial 1 and Trial 4 to establish whether a child improved over the four trials. This may also

test accessibility of the game, and whether children's abilities changed or stayed consistent during the trials.

2.8.3.3. *Initial abstraction score* - this score was based on the D-KEFS 'Initial Abstraction Score' (IAS). It measures the minimum number of items removed from a child's first question across the four trials.

2.8.3.4. *Weighted achievement score* - the 'Weighted Achievement Score' (WAch) was based on the Total Weighted Achievement measure used in the D-KEFS. This score calculates how many questions each child asked in each trial, and measures how close this number is to the optimum number of questions to account for a participant asking too few or too many questions.

The AS, IAS, and WAch scores were used rather than scoring the child based on the total number of questions asked in each trial, as the scores used in this study are less impacted by a child having a 'lucky guess', and instead measure the quality of the questions asked.

2.8.3.5. *Comparison with other measures* - the AS, IAS, and WAch from the Alien Game were compared with the results from the Similarities and Matrix Reasoning tasks using a Spearman's Rank correlation.

2.8.3.6. *Time to first question* - descriptive statistics were recorded on the time taken for the child to answer their first question on each trial. This measure was used to test whether 'mental chronometry', whereby conclusions are drawn about a child's information processing through their reaction time (Meyer, Osman, Irwin, & Yantis, 1988), or 'planfulness', which describes participants tendency to think ahead carefully (Hill et al., 2004), had an effect during this task.

#### 2.8.4. Acceptability

Acceptability of the Alien Game was measured using 5-point Likert scales which also had a visual analogue scale designed for younger children (Appendix H). The results from the Likert scales for the Alien Game, Similarities, and Matrix Reasoning tasks were analysed using descriptive statistics and a Related-

Samples Friedman's Two-Way Analysis of Variance by Ranks to measure whether there was any significant difference between the three Likert scales.

## **2.9. Participants**

### **2.9.1. Demographic Data**

Participants were recruited from three primary schools within or near London. There were 18 participants, 8 males and 10 females, ranging in age from 7-11 years (95 months to 142 months;  $M=125.22$ ,  $SD=14.8$ ). Of these, 16 were primary English speakers and 2 participants had English as an additional language. This needs to be taken into account when generalising the findings from this study.

### **2.9.2. Ability and Cognitive Data**

Table 3 provides non-parametric correlations for the National Curriculum levels for the children in Reading, Writing, and Maths. At the time of conducting this study, 'National Curriculum level' data in schools had changed from numbers, into a scaled score and a judgment on whether a child was 'below the expected level', 'at the expected level', or 'above the expected level' for their age. Even though the children in this study varied in age, this new scoring system enabled a comparison across the children's ability levels. For the purposes of this study, this ordinal data has been transformed into three numbers:

1 = child performing below expected level.

2 = child performing at expected level.

3 = child performing above expected level.

On average, the children's National Curriculum scores indicated these children were performing at the level expected for their age in Reading, Writing, and Maths. From Table 3, we can see that the three types of National Curriculum data were highly correlated ( $p=0.01$ ). Thus, an average of these three scores was

Table 3.

*Non-Parametric Correlations Between and Within National Curriculum Levels for Reading, Writing, and Maths*

	Reading	Writing	Maths
Reading	1		
Writing	.913	1	
Maths	.816	.783	1

calculated to form a mean of participants' overall National Curriculum scores which is presented in Table 4 with the descriptive data.

Table 4 provides descriptive statistics for the participant's scores from the Similarities and Matrix Reasoning subtests, National Curriculum levels and the children's ages (in years) in order to analyse how the participants within this study compare to national norms. The table includes data for the participants' raw scores on the WISC-IV subtests. These scores were also converted into age-scaled scores which are also in Table 4 to allow comparison with published norms (Mean=10, SD=3). From these results, it may be seen that the children scored similarly in their scaled scores for the Similarities and Matrix Reasoning tasks, suggesting that the children's performance was consistent across these two single trial tests of concept formation. However, both of these scores were slightly above the UK average age-scaled score of 10. Additionally, particularly with the Matrix Reasoning task, there was a wide range between the lowest and highest scores achieved by the participants. However, the lowest Similarities scaled score (n=8) fell within the 'normal range' for the child's age group, and thus no children were viewed as having verbal difficulties. The lowest score obtained on the Matrix Reasoning task (n=5) fell within the Low Average range.

Table 4.

*Descriptive Statistics for Age and Key Cognitive Tests*

	Mean	SD	Min.	Max.	Skewness (SE=.536)	Kurtosis (SE=1.038)
Age (years)	9.89	1.28	7	11	-1.100	.108
NatCur (Overall)	2.07	.719	1	3	-.163	-1.172
Matrix Reasoning (Raw Score)	23	4.970	15	34	.559	.458
Matrix Reasoning (Scaled Score)	11.11	3.644	5	19	.326	.143
Similarities (Raw Score)	22.78	4.894	16	32	.534	-.847
Similarities (Scaled Score)	11.44	3.258	8	19	1.148	.634

*Note.* SD = Standard Deviation; SE = Standard Error

All children's data were kept in the study in order to assess how typically developing children with a range of abilities within in a main-stream school performed on the Alien Game. However, the above average WISC-IV scores and the range should be taken into account when generalising the results from this study.

Table 5 displays the measures of association (correlations and effect sizes) between the children's demographic and ability scores. Pearson correlation coefficients ( $\rho$ ) are reported for the children's age and ability data, while etas (effect sizes produced via a one-way ANOVA) are given for associations between test data and gender.

Table 5.

*Measures of Association Between the Demographic and Cognitive Scores*

	Age (rho)	Gender (eta)
Age		.037
National Curriculum	-.191	.296
Similarities Raw Score	-.437	.117
Similarities Scales Score	-.651	.125
Matrix Reasoning Raw Score	.009	.009
Matrix Reasoning Scaled Score	-.300	.001

*Note:* rho = Pearson correlation coefficients; eta = effect sizes produced via a one-way ANOVA.

A significant association between age and similarities raw score was found ( $r(16) = -.651, p = 0.01$ ) with the 7-8 year old children ( $m = 15.25$ ) scoring on average 5 points more than the 9-10 year olds ( $m = 10.43$ ) and 11 year olds ( $m = 10.29$ ). The 7-8 year old children's scores fell within the 'high' range for their age, and thus these scores were above the average range expected for their age group. This finding will be taken into account during further analysis.

It may be seen that gender does not exert a wide influence here. However, there was a moderate association between gender and national curriculum scores,  $F(1,16) = 4.732, \eta = .296, p < 0.05$ . Levene's test for equality of variances was not significant, with the female participants retaining a higher result than the males on average (female = 2.3; male = 1.6).

### **3. RESULTS**

Each child asked fewer than 10 questions to identify the target Alien in each of their four trials. Therefore, each child was deemed as completing each trial.

As this is an exploratory study, different types of analyses were carried out to assess: the types and quality of questions children asked, whether the children adopted strategies, if the Alien Game was enjoyable, and ultimately whether the Alien Game could be used to measure concept formation in children.

#### **3.1. Methods of Analysis**

The procedure for the analysis is described below:

- Questions asked during the Alien Game were coded using a two-phase content analysis of transcripts of each participant's questions:
  - Initial codes were theory driven; based on previous research and neuropsychological testing;
  - A second phase of codes were data driven;
  - The codes were assigned to the data based on this coding scheme derived from the theory and data.
- Descriptive statistics were carried out on the results of the content analysis.
- A second coder rated the data and inter-rater reliability was tested.
- An exploratory content analysis was carried out on the main findings and observations during the study.
- The Alien Game was scored using an Abstraction Score (AS) which was calculated for each trial.
- A Learning Slope (LS) was calculated.
- An Initial Abstraction Score (IAS) was calculated.
- A Weighted Achievement Score (WAch) was calculated.

- The children's average AS, IAS, and WAcH scores were compared to the WISC-IV subtest scores and the children's ages using a Spearman's Rank Correlation.
- Time to first question was assessed.
- Acceptability of the Alien Game was measured using a 5-point Likert scale which assessed a child's enjoyment of the task.

### **3.2. Content Analysis**

Table 6 illustrates the process of generating codes in the two-phase content analysis. In the initial phase, codes were defined based on previous research in this area (see Table 2) and thus were pre-determined by existing scoring systems before collecting data.

The second phase was derived from the data itself, which changed the scoring through both the addition of new codes and some codes from phase 1 being changed or removed. As a new test of concept formation, codes needed to be added or previous codes adapted to create a new scoring system which accommodated the differences in questions asked.

In Table 6, the left hand side column lists the scoring system which was predefined from previous studies. The right hand column lists new codes that were added, changed or removed. As can be seen from the table, an 'Intradimensional' code was added. This was following some of the children in the study asking taxonomical questions (e.g. "Is the Alien yellow? Is it grey?"). The 'Delayed repetition' code was added to capture when a child asked the same question twice, however not immediately afterwards. This was firstly to separate out different types of repetition (whereby the 'Immediate repetition' code captured a question that was repeated twice in a row). If a child repeated the question later on in the set then perhaps it could indicate something other than perseverating.

The 'Medium risk' and 'High risk' strategies were added to capture questions which could either result in removing a large number of pictures with a lucky

Table 6.

*Two Phase Content Analysis*

Phase 1	Phase 2 Added/changed/removed	Added	Changed criteria	Removed
1. 'Constraint seeking'	6. 'Intradimensional'	1. 'Constraint seeking'	4. 'Pseudo-constraint-seeking'.	
2. 'Hypothesis seeking'	7. 'Delayed repetition'	2. 'Hypothesis seeking'		
3. 'Redundant' (renamed as 'Failure to eliminate')	8. 'Medium risk'	5. 'Repetition' (renamed to 'Immediate Repetition')		
4. 'Pseudo-constraint-seeking'	9. 'High risk'			
5. 'Repetition'				

guess, or could result in only removing a few items. It was decided to separate this into two codes to capture the minimum possible items that could be removed from the child's initial question ('High risk' e.g. 4 out of 24 items); whereas the 'Medium risk' code was set to capture questions which removed approximately a quarter of the items (e.g. an initial question which removed 6 out of 24 items). If there were fewer than five items left, then the 'High risk' code was not scored. The 'Medium risk' code could not be given when there were less than 10 Aliens left. The cut-off point was initially set as 5 or less items, but this resulted in an overlap between the "Medium" and "High risk" codes (i.e. both codes could be used to score some of the same questions at that stage).

The criteria for a 'Constraint seeking' question was changed to incorporate a child removing just 1 item when a child has four or less Aliens left. This is because at this stage this strategy was viewed as a child grouping the items into two, similarly sized groups, whereby one of which has to be correct (based on Mosher & Hornsby's, 1966, strategy), rather than being viewed as a 'High risk' strategy.

The criteria for the 'Hypothesis seeking' question was changed as, unlike other 20-Questions tasks, the rules of the Alien Game meant that a child could not ask a question about a single item at the start of the task. Additionally as the 'High risk' code captures a child's strategy to ask a question about a small number of items; the 'Hypothesis seeking' code was redefined to just be coded when a child asked a question which appeared to follow a hypothesis which had no relation to what had previously been asked.

The 'Redundant' code was renamed 'Failure to eliminate' to give a more behavioural description of this type of question.

The 'Pseudo-constraint-seeking' code was removed as this type of question was not deemed possible within the Alien Game.

The coding process resulted in coding scheme which included eight codes (see Table 7). Given the exploratory nature of this study, these codes were exhaustive but were not all mutually exclusive, as a coding unit may correspond to more than

Table 7.

*Coding Scheme*

No.	Coding Category	Description
1.	'Constraint seeking'	Questions which attempt to eliminate two or more Aliens at the time asked (or a question attempting to eliminate one or more Aliens if four or less Aliens remain).
2.	'Hypothesis seeking'	These are questions which are <b>unrelated to the previous question</b> , when there was a <b>possibility of an 'Intradimensional' follow-up</b> to the previous question (see Code 6). For example, if a child asked "Is it blue?" and is told "No", then follows this up by asking "Does it have arms?", despite there being both yellow and grey Aliens left.
3.	'Failure to eliminate'	Questions that did not eliminate any items.
4.	'Delayed repetition'	A repeated a question from earlier in that trial.
5.	'Immediate repetition'	Preservation; repeating the question just before it.
6.	'Intradimensional'	Asking a question related to the question immediately before it. Therefore, following the same taxonomy (e.g. "Is the Alien a triangle?" "Is the Alien a circle?").
7.	'Medium risk'	A question that would at a minimum eliminate a quarter or less of the options (see Appendix K for scoring sheet). This is coded whether or not the question yielded a correct answer. This Code cannot be applied if there are only 10 or less Aliens left.
8.	'High risk'	A question which at a minimum eliminates a sixth or less of the items left (see Appendix K for scoring sheet), regardless of whether the questions yielded correct responses or not. Do not code when there are fewer than five Aliens left.

one code simultaneously. Codes that could overlap were included in order to investigate which codes were the most appropriate to determine patterns of behaviour from this data.

### 3.2.1. Quantitative Analysis

When analysing the types of questions asked, the children in this study appeared to follow a somewhat similar trend (see Table 8). Of all the questions asked, the most commonly used type of question was 'Constraint Seeking', with every child asking at least one question that was given this code in 100% of the trials. Thirteen out of 18 of the children used a question coded as 'Intradimensional', which was the same amount of children who used a 'Hypothesis seeking' strategy, except those who used an 'Intradimensional' question were more likely to use these frequently, with 10 out of the 13 children asking at least one of these questions on 50% of their trials.

Although only 8.31% of the questions were classified as 'High risk', 12 out of 18 children asked a question that was given this code at least once. Likewise, although only 3.88% of questions were coded as 'Medium risk', 8 out of 18 children asked a question that was given this code. Therefore if the children were going to ask a question that could potentially eliminate fewer items, they were more likely to ask a question that could potentially eliminate more items if they had a 'lucky guess'. Of the children that asked a 'High risk' question, 50% did so on more than one trial, and 2 out of the 8 children that asked a 'Medium risk' question did so on more than one of their trials. Five out of 18 children did not ask either a 'Medium risk' or 'High risk' question. Six children asked a 'Failed to eliminate' question; two of which asked these questions on more than one trial. Only one child asked a 'Delayed repetition' question once. None of the questions were coded as 'Immediate repetition'. A similar pattern can be seen within each age groups within Table 8.

### 3.2.2. Inter-Rater Reliability

After the researcher coded the data, a second coder rated 10% of the data. The two sets of codes were then compared using a Cohen's Kappa in order to

Table 8.

*Percentage of Codes Used and Number of Questions Asked Overall*

	Overall (%)	Ages 7-8 (%)	Ages 9-10 (%)	Ages 11 (%)
'Constraint seeking'	92.8	93.42	95.07	90.21
'Hypothesis seeking'	8.03	7.89	9.15	7
'Failure to eliminate'	2.5	1.32	1.41	4.2
'Delayed repetition'	0.28	-	-	-
'Immediate repetition'	-	-	-	-
'Intradimensional'	9.7	14.47	7.75	9.09
'Medium risk'	3.88	-	3.52	6.29
'High risk'	8.31	9.21	7.75	8.39
Number of questions asked overall	20.11	18	20.29	21.14
Mean number of questions per trial	5.1	4.81	5.07	5.29

determine if there was agreement between the two coders. An 'almost perfect' agreement (Viera & Garrett, 2005) was found  $k=.945$   $p= <001$  (standard error=.052).

### 3.2.3. Exploratory Content Analysis

It was noted that all children used full and complete sentences when asking questions, for example asking "Does your alien have blue skin?", rather than just saying "blue". There is a possibility that this might be different when conducting this test on clinically impaired children.

The vast majority of questions were coded as 'Constraint seeking'. This indicates the children's ability to survey the scene in front of them and group the Aliens' characteristics together into groups that were usually large whereby whether or not their question yielded a correct response, a group of Aliens would be eliminated (e.g. if a child asked "Is it yellow?" for their first question, that question would eliminate either 7 or 17 items depending on whether the answer was 'yes' or 'no'). Additionally, this indicates the children were able to hold in mind the task rule of guessing the correct answer in as few questions as possible. It was interesting to note that although there were two questions which eliminated half of the items ("Does it have a tail?"; "Does it have no legs?") only one of these questions was asked just once as the first question on a trial. It is not clear why these questions were not asked frequently. There is a chance that this was due to the tail and legs being a less prominent feature of the aliens (compared with some of their other characteristics such as their skin colour or shape). Perhaps it was because these questions did not offer a possibility of a follow-up, 'Intradimensional' question, or perhaps it was because these features were not viewed by the children as the main features of the Aliens (as opposed to colour or shape), or whether, when given the choice, children preferred to ask a question whereby they could eliminate two thirds of the answers, rather than half.

The second most commonly used type of questions were 'Intradimensional'. As children used more 'Intradimensional' questions than 'Hypothesis seeking', this indicates this group of children were slightly more likely to relate their questions to what they had previously asked, than to switch to a question about an

unrelated, new characteristic. The 'Intradimensional' questions indicated a child's strategy of trying to eliminate groups of items, based on a taxonomical aspect. The perceptual taxonomies that children used in these instances included trying to ascertain what colour, shape, or number of body parts (e.g. eyes or legs) an alien had. The children were not asked why they chose their questions, but some children verbalised thoughts in between asking questions, such as one child who stated "So, it's easier to go with colour" before choosing the correct colour on his first go. It is therefore possible that many of the children's 'lucky guesses', (for example, one child guessed the correct colour of the target Alien in his first question of each trial) could have been followed up with an 'Intradimensional' question if the child had not guessed the correct aspect initially. Thus, perhaps the percentage usage of this type of question might be under-representative of the intent to use an 'Intradimensional', taxonomical structure in their questioning. The use of these taxonomical structures could perhaps also be seen in many children using the same or a similar first question on the first question of each task, for example one child asked "Is your alien blue?" for the first three tasks. On the first two trials this yielded an incorrect response, and the child followed this question with an 'Intradimensional' question about another colour. On her third task, this question yielded a correct answer, and the child began her final trial by asking "Is this alien yellow?".

When using 'Intradimensional' questions, it is important that once a child has successfully identified the correct aspect of this taxonomy that they are able to change set and focus on a different, perceptual aspect of the pictures. The majority of children in this study did that on most questions; those that were not able to change set then asked a question that was coded as 'Failure to eliminate'. 'Failure to eliminate' questions were also asked when a child was not following a taxonomical structure (e.g. asking "Does the Alien have wings?" when only winged aliens remained). It is not clear why these questions were asked. Perhaps the child could not remember the previous question that they had asked, or they had not surveyed the scene, and so did not notice that their question was redundant. Of the children that asked 'Failure to eliminate' questions; all but one commented afterwards on their mistake, and the other child sighed loudly, which was interpreted as him realising he had made a mistake. One of these children

was the only child who said that he was unfamiliar with the 20-Questions style games (including having not played a game such as 'Guess Who?'). It is interesting to note that typically developing children do make these mistakes, but that with these particular children, they all noticed afterwards. However, it is not clear whether the child would have noticed these mistakes in the original 20-Questions tasks, or whether these mistakes were unambiguous due to the added element of being able to eliminate items on the hinged frame, which the child was thus unable to do after asking a redundant question.

It is also important to note that within a typically developing sample, some children will ask questions coded as 'High risk' or 'Medium risk'. It is not clear why the children used these strategies; whether they should be viewed as *errors*, for example as a result of a child not surveying the scene thoroughly, or whether this strategy was intentional, and could be useful in some instances, as when this strategy pays off, it could result in a child needing fewer questions overall.

Only one child asked a 'Delayed repetition' question, which could indicate that this child had difficulty in remembering what questions had already been asked. This type of question may be more common in a clinical sample (e.g. Baldo et al., 2004).

Although the children were told that they were not able to ask a question about an Alien's name, two children did break this rule.

### **3.3. Other Measures**

Alongside the content analysis, measures based previous studies were incorporated. These were:

1. Abstraction Score
2. Learning Slope
3. Initial Abstraction Score
4. Weighted Achievement Score
5. Comparison with WISC-IV Matrix Reasoning and Similarities tasks.
6. Time to First Question

### 3.3.1. Abstraction Score

Table 9 displays the mean 'Abstraction Scores' (AS). Different approaches to calculating an AS have been used in different studies. Within this study, the AS was calculated by adding up the minimum number of items eliminated from a child's first three questions across their four trials (e.g. if 9 characters remained and a child asked a question about 3 of them, their abstraction score for that question would be 3, regardless of whether the question yielded a correct, 3, or incorrect, 6, answer.).

Table 9.

#### *Mean Abstraction Scores*

	Trial	AS	Min.	Max.	SD
Overall	1	15.50	10	19	2.895
	2	15.11	9	20	4.143
	3	14.33	10	19	2.114
	4	14.78	7	19	2.942
7-8 year olds	1	15.00	12	18	2.944
	2	15.00	10	20	5.228
	3	13.50	11	15	1.732
	4	15.75	14	17	1.258
9-10 year olds	1	15.43	10	19	3.690
	2	15.71	9	19	4.309
	3	16.00	14	19	1.633
	4	14.29	7	19	4.112
11 year olds	1	15.86	13	19	2.340
	2	14.57	10	19	3.952
	3	13.14	10	16	1.773
	4	14.71	12	18	2.430

The minimum number removed was used in the calculation rather than the maximum, in order to assess the quality of the children’s questions, and to prevent a child from gaining a high score due to a lucky guess.

Overall, the average AS was 14.87 (min=12.5, max=17.75). Table 9 also shows the average AS per age group across the four trials. It was noted that the two children with the lowest AS for a single trial (n=7, 9) asked ‘High risk’ questions within these trials. The person with the highest AS (n=20) asked questions that were all coded as ‘Constraint seeking’ questions, with one also coded as ‘Intradimensional’.

### 3.3.2. Learning Slope

A Learning Slope (LS) was calculated by subtracting children’s AS in Trial 1 from their AS in Trial 4. Table 10 shows the calculations and results for the mean LS overall and across age groups (for individual LS scores see Appendix M). As can be seen from Table 10, the mean LS was -0.72, meaning that on average, the children’s mean score at Trial 4 was slightly lower than their score at Trial 1. No Learning Slope was observed.

The one child who was unfamiliar with 20-Question style games had an LS of 3, demonstrating that he improved from his first to his fourth trial.

Table 10.

#### *Mean Learning Slope Scores*

Age	Trial 4 - Trial 1		= Learning Slope	Min.	Max.	SD
Overall	4.78	15.50	-0.72	-9	4	3.404
7-8 years	15.75	15.00	0.75	-1	3	2.062
9-10 years	14.29	15.43	-1.14	-9	4	4.018
11 years	14.71	15.86	-1.15	-7	3	3.450

### 3.3.3. Initial Abstraction Score

Table 11 shows each participant's Initial Abstraction Score (IAS) and the scores overall. This was calculated from the lowest abstraction obtained from the child's first question in each of the four trials. The maximum IAS was 12, which was 50% of the items (a score that could be obtained by only two questions). This score was only obtained once. The lowest possible score obtained was 4 which was a sixth of all items (a score that could be obtained by one question). This score was given three times. The mean IAS was 31.22 ( $m=7.81$  for individual trials).

Table 11.

#### *Initial Abstraction Score*

Participants	Trial 1	Trial 2	Trial 3	Trial 4	Total
1	8	7	7	11	33
2	9	9	9	7	34
3	8	7	9	8	32
4	10	10	10	7	37
5	9	9	9	9	36
6	6	6	6	6	24
7	7	12	11	7	37
8	4	9	9	5	27
9	9	9	9	7	34
10	9	9	4	7	29
11	6	7	8	11	32
12	9	9	8	7	33
13	9	8	6	7	30
14	8	4	7	6	25
15	6	7	9	9	31
16	7	7	9	8	31
17	8	7	9	7	31
18	9	4	5	8	26
Total	141	140	144	137	562
Average	7.83	7.78	8	7.61	31.22 (7.81)

### 3.3.4. Weighted Achievement Score

The scoring structure for deriving a child's 'Weighted Achievement Score' (WAch) was based on the approach used to score the D-KEFS 20-Questions task (see Table 12). The scoring system measures whether the child used an optimum number of questions in order to identify the target item, with a lower score being given if the target Alien was guessed within too few or too many questions (See Table 12). The optimum score of '5' was derived from this being the average score obtained by the children. As can be seen in Table 13 the mean overall score was 16.28 (m=4.07 for an individual trial).

Table 12.

#### *Weighted Achievement Scoring System*

Number of Questions	Weighted Achievement Score
1-2	1
3	2
4	4
5	5
6	4
7	3
8	2
9-10	1
>10	0

Table 13.

*Weighted Achievement Score*

Participant	Trial 1	Trial 2	Trial 3	Trial 4	Overall
1	4	5	2	5	16
2	5	5	5	5	20
3	4	4	5	4	17
4	3	5	4	5	17
5	4	5	4	4	17
6	3	4	4	4	15
7	5	4	2	4	15
8	4	4	5	4	17
9	4	4	4	5	17
10	4	4	3	5	16
11	4	4	5	3	16
12	3	4	5	5	17
13	4	3	2	3	12
14	5	4	2	5	16
15	1	4	4	4	13
16	4	4	4	5	17
17	5	5	5	5	20
18	4	5	4	2	15
<b>Total</b>	70	77	69	77	293
<b>Average</b>	3.89	4.28	3.83	4.28	16.28 (4.07)

3.3.5. Comparison with Other Measures

WISC-IV Similarities and Matrix Reasoning tests were scored according to their test criteria, which included age-scaled scores being calculated (see Table 4). The AS, IAS, and WACH results were correlated with each other and the results from the Similarities and Matrix Reasoning tasks using a Spearman's rank correlation in order to test for convergent validity (Table 14). Both the raw scores and scaled scores were calculated for the WISC-IV subtests as the AS has not

Table 14.

*Spearman's Rank Correlation Comparing Abstraction Score with Similarities and Matrix Reasoning*

		Age	WAch	AS	IAS	Similarities Raw Score	Similarities Scaled Score	Matrix Reasoning Raw Score	Matrix Reasoning Scaled Score
Age	Coefficient	1.000							
	Sig.								
WAch	Coefficient	-.259	1.000						
	Sig.	.299							
AS	Coefficient	-.089	.146	1.000					
	Sig.	.725	.564						
IAS	Coefficient	-.515	.390	.331	1.000				
	Sig.	.029	.109	.180					
Similarities Raw Score	Coefficient	-.292	.213	-.073	-.008	1.000			
	Sig.	.240	.395	.773	.976				
Similarities Scaled Score	Coefficient	-.612	.247	.056	.255	.890	1.000		
	Sig.	.007	.323	.826	.307	.000			
Matrix Reasoning Raw Score	Coefficient	-.043	-.323	.161	-.218	.305	.301	1.000	
	Sig.	.865	.191	.524	.385	.218	.224		
Matrix Reasoning Scaled Score	Coefficient	-.358	-.134	.150	.064	.437	.559	.882	1.000
	Sig.	.145	.596	.552	.800	.070	.016	.000	

been age-controlled at this stage. As can be seen Table 14 when comparing the Alien Game measures, the IAS and WACH scores had a moderate positive relationship with each other ( $r=.390$ ,  $p=.109$ ), as did the IAS and the AS ( $r=.331$ ,  $p=.180$ ). However, a weak correlation was found between the WACH and AS ( $r=.146$ ,  $p=.564$ )

A very small correlation can be seen between the AS and the results from the Similarities and Matrix Reasoning. A positive medium correlation found between the children's IAS score and Similarities scaled scores. Similarly a positive medium correlation was found between the children's WACH and Similarities raw and scaled scores. This indicates that a positive relationship was found between the children's WACH and IAS scores with their Similarities score. A negative medium correlation was found between the children's IAS scores and their Matrix Reasoning raw score. A negative medium correlation was also found with the children's WACH scores and their Matrix Reasoning raw score, indicating that a higher WACH score correlated with a lower Matrix Reasoning score.

When comparing the Alien Game scores with age (in months), no correlation was found between AS and age ( $r=-.089$ ,  $p=.725$ ), and a weak correlation was found between WACH and age ( $r=-.259$ ,  $p=.299$ ). However, a significant correlation was found between IAS and age ( $r=-.515$ ,  $p<.05$ ), whereby the younger children had a significantly better score than the older children. This result can be seen as being in line with the fact that the younger children scored higher on the WISC-IV subtests.

### 3.3.6. Time to First Question

The time to the first question across the four trials was recorded (and rounded up or down to the nearest half a second). After the children were prompted to ask their first question, the longest a child took before asking was 6 seconds, and the quickest was less than half a second ( $m= 2.29$  seconds;  $SD 1.14$  seconds).

Three children took less than half a second to respond on one of their trials, and all of their first questions were coded as 'Constraint seeking'. The IAS for these trials were 7, 8, and 10. The child who took 6 second to ask a question in one of his trials asked a 'Constraint seeking' question, and obtained the highest possible

IAS of 12. On all other trials, this child took 1 or 2 seconds to ask a question. The child who took 6 second to respond on one trial had an overall AS of 13.75 and a WAcH of 4. The AS from those who spent less than half a second to ask their first question had overall AS scores of 13, 16.75 and 17.75. Their WAcH scores were 4, 4, and 5.

### 3.3.7. Acceptability

The 5-point Likert scales for the Alien Game, Similarities, and Matrix Reasoning tasks were compared using a Friedman’s Two-Way Analysis by Ranks (See Table 15 and Figure 2). Here it can be seen that the Alien Game obtained the highest score of the three tasks, which was significantly higher than the results for the Similarities task ( $X^2(2)=1.028$ ,  $p=.002$ ).

Table 15.

*Friedman’s Two-Way Analysis By Ranks*

	Test Statistic	Std. Error	Std. Test Statistic	Sig.
Similarities-Alien	1.028	.333	3.083	.002
Matrices-Alien	.306	.333	.917	.359
Similarities-Matrices	-.722	.333	-2.167	.030

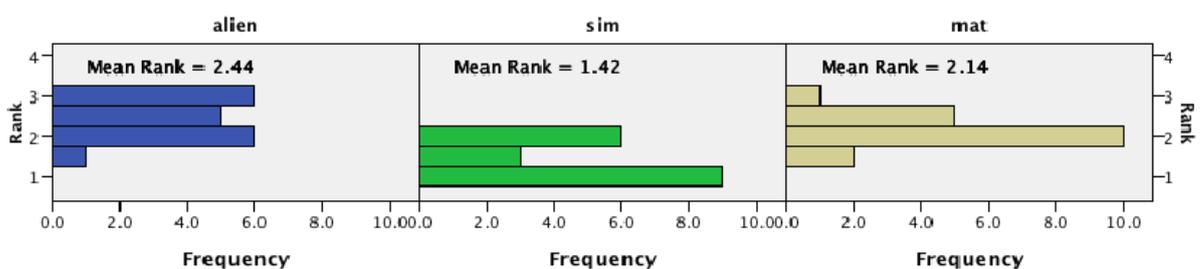


Figure 2. *Related-Samples Friedman’s Two-Way Analysis of Variance by Ranks.*

Note. alien = Alien Game, sim = Similarities task, mat = Matrix Reasoning.

## **4. DISCUSSION**

This section will re-orientate the reader to the research questions before providing a summary of the research findings. It will then revisit the previous literature explored in the introduction on testing children's concept formation abilities before critically reviewing this study's findings. It will finish by drawing conclusions and clinical implications from this study before exploring possible directions for future research.

### **4.1. Revisiting the Research Questions**

The aim of this present study was to examine whether a game-like procedure could be used as a standardised test of concept formation with typically developing children. This study built on ideas proposed by Mosher and Hornsby (1966) and Alderson-Day and McGonigle-Chalmers (2011) to create a 20-Questions task in the form of a game; the Alien Game. This study aimed to address three questions:

- 1) Do children (from a non-clinical sample) adopt similar strategies when completing the Alien Game?
  
- 2) Can we establish normative performance characteristics (including scores and typical patterns of responding) from the results, that capture normal variation in concept formation?
  
- 3) Do children engage better with the Alien Game than they do with other standardised measures of concept formation?

#### **4.1.1. Research Question 1**

This study aimed to investigate whether children from a non-clinical sample adopted similar strategies when completing a new task; the Alien Game. This was analysed through a content analysis. The children's questions were

orthographically transcribed. A coding system was then generated through a two-phase content analysis created from previous tests of concept formation and from the children's responses in the Alien Game. Responses were coded using this system, before quantitative and qualitative analysis was carried out on this data.

4.1.1.1. *Interpretation of findings:* the results of the content analysis indicated that children in this sample generally did adopt similar strategies while playing the Alien Game. Firstly, each child gave rich answers indicating that they were able to give voice to their reasons and communicate their answers with clarity. Therefore, the children answered with some confidence in their answers, rather than responding in partial or incomplete answers which could indicate that these were made-up or ad-hoc answers. Within every trial, the vast majority of questions were 'Constraint seeking' and aimed to eliminate more than two items by grouping together larger groups of characteristics. This meant that whether the question yielded a right or wrong answer, a large number of non-target Aliens would be eliminated (e.g. initially asking "Does it have legs?" eliminates 12 items, whether or not the question yielded a correct answer). Each child asked at least one 'Constraint seeking' question within each trial.

When there was an opportunity to ask a question related to the child's previous question, it was found that the children asked a related, 'Intradimensional' question slightly more frequently than an unrelated, 'Hypothesis seeking' question. The same number of children asked 'Intradimensional' and 'Hypothesis seeking' questions (n=13), but those who asked 'Intradimensional' questions asked these more frequently. The children who asked 'Intradimensional' questions can be viewed as adopting a taxonomical structure to their questioning. However, the percentage of 'Intradimensional' questions asked may be unrepresentative of the children's intent to adopt a taxonomical structure, as the children who guessed the correct aspect of the taxonomy with their first question would have asked a 'Failure to eliminate' question if they continued to ask questions about that particular taxonomy; thus 'Intradimensional' questioning was not possible at all stages of the game.

Although only a few questions were coded as 'High risk', it was found that most children asked at least one 'High risk' question in at least one trial; many asked this type of question in more than one trial. Similarly, although few questions were coded as 'Medium risk', many children asked at least one 'Medium risk' question. Therefore, these questions were not unusual within a typically developing sample. However, these questions were consistently asked alongside a majority of questions which were not coded as a 'risk' question.

Some children asked a question that resulted in a 'Failure to eliminate' code. Each time this occurred, the child appeared to recognise their mistake. Of these children, a few asked this type of question in more than one trial. No child asked a question coded as 'Immediate repetition' (perseveration), and only one child asked one question that was coded as 'Delayed repetition' (falling off set). Given the exploratory nature of this research study, these codes were exhaustive but not all mutually exclusive, as it was acknowledged that a coding unit may correspond to more than one code simultaneously. This allowed for a more thorough exploration of the strategies used by the children.

Thus, the children within this study adopted similar strategies which involved using a majority of 'Constraint seeking' questions. Most children also asked at least one question which involved some risk but could result in a 'high reward' if the child had a lucky guess. When a possible related question was available, most children asked at least one question which related to their previous question; the same proportion of children also asked at least one question that was unrelated to their previous question, despite the available related question.

From the strategies observed in this sample it would be expected that typically developing children are very unlikely to repeat a question within the same trial. If a typically developing child asked a question which did not eliminate any items, it would be expected that they would notice their mistake.

#### 4.1.2. Research Question 2

The second research question asked whether it was possible to establish normative performance characteristics from the results that captured normal

variation of concept formation, including scores and typical patterns of responding. In order to do this, the Alien Game was scored using several different measures to test which most accurately described the children's responses on the game. In order to complete the task, the children had to identify the correct Alien using 10 or fewer questions, which all of the children did. The scores were: an Abstraction Score (AS), a Learning Slope (LS), an Initial Abstraction Score (IAS), and a Weighted Achievement Score (WAch). The Time to First Question was also measured. Finally, the AS, IAS, and WAch scores from the Alien Game were correlated with the Similarities and Matrix Reasoning scores. This was to assess whether the children's verbal and visuo-spatial abilities correlated with their scores on the Alien Game, by using a verbal and non-verbal single trial test of concept formation. The children's typical patterns of responding on the Alien Game were assessed in the content analysis (see section 4.1.1.1.).

4.1.2.1. *Interpretation of the findings:* within individual trials, the children who obtained the lowest AS used 'High risk' questions within that particular trial, whereas the child who obtained the highest AS only used questions coded as 'Constraint seeking'.

The LS scores showed that overall the children's results remained fairly consistent throughout. This indicates that this task remained a consistent measure of a child's concept formation abilities throughout and was not affected by practice effects. However, the only child who was not familiar with 20-Questions style tasks had one of the highest LS scores, and therefore did improve. It was additionally observed that he asked two 'Failure to eliminate' questions. Thus, he may have benefited from having a practice trial before the task, which would have been possible due to performance on this task not appearing to be affected by practice affects. However, it is equally important to note that all other children who had a positive LS were familiar with 20-Questions style tasks. Additionally, this task used novel pictures so that no child would be more familiar with this exact task.

The IAS measured the minimum possible number of items removed by the child's first question on each trial. The results from individual trials ranged from the lowest to the highest possible score. However, no child scored the lowest or highest score on more than one trial. The child who obtained the highest score asked a 'Constraint seeking' question. The children who obtained the lowest score asked a question coded as 'High risk'. The norm was asking a question that removed approximately a third of the answers at a minimum.

The WACH score provided a weighted score based on the optimum number of questions asked in each trial. This accounted for children having a 'lucky guess', or using a strategy which required a greater number of questions. It was found that typically developing children scored just below the 'optimum' score on average. This could have been affected by the children's use of at least one risk question. The results from the AS, IAS, and WACH indicate that the best strategy for obtaining a high score is to use 'Constraint seeking' questions.

For Time to Respond, overall children took between half a second and six seconds to ask their first question. It was noted that the child who took the longest time to respond asked an initial question that eliminated half of the items, however, his overall AS and WACH scores were the same or worse than other children who spent less than half a second to respond. Thus, it is not clear whether it is best for a child to spend a longer or shorter time thinking about their answer. The difficulty with this measure is that we cannot accurately measure a child's time spent thinking about their question, as they may have already thought of it before they were prompted to ask due to the Aliens being visible before the child started the task and the same set of Aliens being used across all four trials. Thus, it was difficult to establish whether 'planfulness' and carefully thinking ahead (Hill et al., 2004) occurred if a child took longer to answer a question, or whether this served as a measure of mental chronometry, whereby a faster response time may indicate a person being more alert (Meyer et al., 1988). Some children might have taken longer to answer due to being shy, cautious, or engaging in an off task activity. Alternatively, a child may instantly grasp what was required of them and answer quickly, without this being an indicator that they had not thought their answer through. As an exploratory study it was important to

try out different measures. However, from these inconclusive findings, the Time to First Question measure did not appear to have the potential to become a measure of normative performance.

When comparing the AS, IAS, and WACH scores for the Alien Game, the IAS was moderately correlated with both the WACH and AS, while a weak correlation was found between the WACH and AS, indicating that these two measures were testing slightly different things and might map onto tap into different performances in real life, or be similar to different instruments.

When comparing the Alien Game scores to the Similarities and Matrix Reasoning tests, the children's IAS and WACH scores had a positive, medium correlation with their Similarities scaled scores, with the WACH also being positively correlated with the Similarities raw score. Whereas both of these scores also had a medium, negative correlation with the Matrix Reasoning raw score. It is not clear why a negative correlation was found. However, these scores indicate that the children's verbal abilities were more predictive of a higher IAS and WACH score and thus these scores appeared to be based more on a child's verbal than visuo-spatial abilities. No strong correlation was found between any of the Alien Game scores and either WISC-IV sub-tests. This suggests that the Alien Game scores measure somewhat different aspects of concept formation to the WISC-IV subtests, which may be due to the Alien Game being a multi-trial task involving both verbal and non-verbal abilities. The multi-trial format meant that the Alien Game was more dependent on working memory. This is because children had to keep track of questions they had previously asked while following a self-ordered hypothesis formation, making this task qualitatively different. Although the WISC-IV sub-tests were mainly used in this study as measures of verbal and visuo-spatial abilities, they also provided a useful comparison of children's concept formation without leading to participants experiencing practice effects as these are both single trial tests.

In summary, the normative performance characteristics were captured using an AS, IAS, and WACH. These all provided measures which were not impacted by a child having a 'lucky guess'. The best way to get a high score on any of these

measures was by asking 'Constraint seeking' questions, which all the children in this study did. The LS indicated that the children's performance remained fairly consistent between Trials 1-4. The Time to First Question measure did not appear to be a useful measure of normative performance.

#### 4.1.3. Research Question 3

The final question asked whether children engaged better with the Alien Game than they did with other standardised measures of concept formation. This was tested by comparing Likert scales which measured enjoyment of the Alien Game, Similarities, and Matrix Reasoning tasks.

4.1.3.1. *Interpretation of the findings:* the children enjoyed the Alien Game significantly more than the Similarities task, and somewhat more than the Matrix Reasoning tasks. The children were observed to engage well throughout the task. From these results, it can be concluded that the children found the Alien Game to be more enjoyable and engaging than the other tasks.

## 4.2. Summary of Findings

This study tested the utility of a new neuropsychological tool based on the '20-Questions' task. The Alien Game used novel pictures of 'Aliens' so that no child would be more familiar with the pictures. Every child completed the game (by asking less than 10 questions), and appeared to understand the rules. This game was specifically made for children. The children were observed to engage well with the game and appeared to try their best. The children liked this task more than other measures. A norm from this study was that the children were observed to give rich answers and to be able to communicate their responses with clarity, including the younger children.

Normative performance was established in this sample through coding the questions asked and in scoring an AS, IAS, and WACH score. Children were found to tend to use 'Constraint seeking' questions and occasionally take risks.

### 4.3. Evaluating the Present Study in Relation to Previous Research

Concept formation is a cognitive process which enables someone to categorise objects or experiences under certain rules or classes (Hunt, 1962). These abilities are important for children in developing their academic and adaptive skills (Gligorović & Buha, 2013). Yet, very few clinical tests have been designed to specifically test concept formation in children (Channon & Crawford, 1999), despite research indicating that children and adults are engaged by different types of testing and may differ in the way they form concepts. Some previous tasks also had instructions which were too complex, or a research design which was not sensitive enough, leading them to underestimate or not accurately assess a child's concept formation abilities (Wood et al., 2001).

Additionally, the majority of current tests of concept formation for children and adults are single trial tests which can be viewed as lacking ecological validity as they cannot assess a child's ability to problem solve through multiple questions.

This study aimed to build on previous research and tests of concept formation, including the work by Mosher and Hornsby (1966) and Alderson-Day and McGonigle-Chalmers (2011), to create:

- A multi-trial test of concept formation specifically designed for children which can be used with younger children.
- A new, novel test based on the '20-Questions' task with the aim of establishing normative performance characteristics (scores and typical patterns of responding) in a typically-developing sample which could be used as a standardised neuropsychological test of concept formation.
- A game-like structure through the use of a hinged board and in using characters with a similar amount of characteristics to those used in 'Guess Who?' to introduce an element of chance and thus enhance the 'game-like' format.
- An enjoyable game which engaged children.

- A task with easy to understand instructions, and which is sensitive enough to detect children's concept formation abilities and difficulties.

Steps were taken to make the test specifically designed for children. This included using a game-like procedure (Gioia, 2015) with instructions that could be easily understood by children. Additionally, the task was designed to be suitable to the nature of impairments that children with frontal lobe difficulties may experience (Taylor & Fletcher, 1990) by making the task short to help children maintain attention (Marton, 2008) and in allowing physical elimination of items to reduce working memory demands (Alderson-Day & McGonigle-Chalmers, 2011). In line with Gioia's (2015) finding that using a game-like procedure can increase engagement in children, the participants preferred the Alien Game task to two tasks which are similar to those used with adults in the WAIS-IV. The children all completed the task. However, this could indicate that the children found this task easy and that, like Alderson-Day & McGonigle-Chalmers' (2011) novel task, this task experienced a 'ceiling effect'. Alternatively, it could mean that we should expect typically developing children to complete this task. This will need to be tested in future research with other groups of children who have difficulties in concept formation (see section 4.7.).

This task used novel stimuli, yet as nearly all of the children were familiar with 20-Questions tasks, there is a chance that some children also used their prior knowledge gained from playing similar tasks before to inform their decisions on the best type of question to ask. The ability for children to form category selections based on prior knowledge of perceptual similarities and differences allows for the organisation of novel information in a manageable way (Alderson-Day & McGonigle-Chalmers, 2011). Previous research shows that a combination of prior knowledge and novel situation are important in category formation. The one child who had not played a 20-Questions task before potentially had to rely more on his ability to develop concepts in novel situations than the other children. Two of his questions were coded as 'Failure to eliminate', after which he recognised his mistakes. Previous studies which used a 20-Questions task also asked participants if they were familiar with 20-Questions style tasks, with some

offering a practice to participants (e.g. Alderson-Day & McGonigle-Chalmers, 2011).

When comparing this task to traditional 20-Questions tasks (e.g. Mosher & Hornsby, 1966), the obvious difference with the present task is the game-like structure using 24 novel characters on a hinged board; a strategy also adopted by Alderson-Day and McGonigle-Chalmers (2011). This present task was similar to Alderson-Day and McGonigle-Chalmers' task in that children were able to physically eliminate items, and the children were only permitted to ask 10 questions to identify the target alien. 'Aliens' were chosen so that no child was more familiar with the pictures in order to make this task more culturally fair. The differences are this present task's characteristics were based on a game-like structure rather than distribution of characteristics being mainly being split into half or a quarter of the items. Also, this study was designed with the aim of capturing data that had the potential to produce norms in future testing with a larger and representative sample.

Consistent with findings from the typically developing participant's questions in other 20-Questions tasks (e.g. Alderson-Day & McGonigle-Chalmers, 2011; Baldo et al., 2004; Mosher & Hornsby, 1966), the children in this study asked a majority of 'Constraint seeking' questions. Although all studies used a slightly different criteria to describe a 'Constraint seeking' question, the consistent component is an attempt to remove large groups of items regardless of whether the question yields a correct answer. Some previous studies defined 'Constraint seeking' as a question which attempted to eliminate half of the options available. In those studies, the distribution of characteristics was set so that certain questions could potentially eliminate half of the items. This study adopted the definition of 'Constraint seeking' as a question which referred to two or more items. The more varied distribution of item characteristics within this study meant that there were only two questions which could possibly eliminate half of the items at the start ("Does your alien have no legs?" "Does your alien have a tail?"). Only one of these questions was asked once as an initial question; children more typically opted for questions regarding colour or shape for their first question.

Traditional 20-Questions tests only scored for 'Constraint seeking', 'Hypothesis seeking', and 'Pseudo-constraint seeking' questions. The game-like structure of the present study, led to the addition of codes which measured 'risk'. The coding of such questions added an extra dimension to assessing the children's strategies. Most of the children were observed to use a risk strategy at least once.

Similarly to previous tests, 'Hypothesis seeking' questions were coded. However, the criteria was adopted to suit this study. Previous studies that recorded 'Hypothesis seeking' or 'Hypothesis scanning' questions often used this code to capture questions which only eliminated one item, and thus could have potentially led to a person guessing a correct answer in one question (e.g. "Is it the cat?"), or to capture questions that did not follow a taxonomy. Within this study, a 'Hypothesis seeking' code was given when there was a possibility of a follow-up, taxonomical question, which the participant did not ask. It did not refer to how many items the question eliminated. In contrast with other 20-Questions tasks, there were no questions that eliminated just one item at the beginning of each trial. At the stages of the task when children were able to ask a question about just one item, these questions were coded as either 'High risk' or 'Constraint seeking' depending on the number of items left.

In contrast to the 'Hypothesis seeking' code, the 'Intradimensional' code was given when children used a taxonomical question related to their previous question. As the youngest child in this study was 7 years old; prior research indicated that these children would therefore be able to use taxonomic relations and can shift from one grouping principle to another (Reichard et al., 1943). This prior research was supported by this study as over two thirds of children asked a taxonomic question and all the children were able to switch between grouping principles.

Questions that were not coded as 'Constraint seeking' were either coded as 'Medium risk', 'High risk', 'Failure to eliminate', or as a type of repeated question. A few children asked a question coded as 'Failure to eliminate', and typically noticed afterwards. Although this was rare in this sample, previous research

suggests that children with impaired concept formation abilities may ask more 'Failure to eliminate' questions as perseveration might occur with some children who have impairments in working memory. Therefore, it is recommended that this code be kept in for future research studies (see section 4.7.).

This study only used a sample of typically developing children at this exploratory stage. Previous research using 20-Questions style tasks have found that different clinical groups adopt different strategies on these tasks than controls. Alderson-Day and McGonigle-Chalmers (2011) found that children with ASD asked more questions than controls to complete the game. In other words, they asked questions which eliminated fewer items at a time (i.e. more 'Hypothesis seeking' questions than 'Constraint seeking' where 'Hypothesis seeking' questions only eliminated one item). Baldo et al. (2004) found that patients with focal prefrontal lesions asked significantly more questions than controls on the 20-Questions task. Future research will need to test the Alien Game with people with concept formation difficulties (see section 4.7.). However, it is important to note that from this study's results, if a child does use a risk strategy, this is not necessarily an indicator of a pathological problem, as asking one or two of these types of questions across the four trials, among a majority of 'Constraint seeking' questions, is a relatively common occurrence within a typically developing sample.

When comparing the scoring system, different 20-Questions studies used a different criteria for their AS, including Baldo et al. (2004) who used the first four questions asked across trials 1-4 in their scoring. Within this study, the minimum number of items removed from the first three questions asked in each trial was used because some children completed a trial with only three questions.

Similarly to the D-KEFS scoring system, the IAS and WACH were scored. Thus, this study adopted the same strategy taken by the D-KEFS, and the IAS was created from a sum of the fewest number of items eliminated. Where this present study's scoring system differs to the D-KEFS is in also scoring a person's AS.

## 4.4. Critical Evaluation

### 4.4.1. Strengths of the Current Study

4.4.1.1. *Design* - this current research is one of the first studies to explore the utility of a test of concept formation designed specifically for children with the aim of standardising it as an assessment tool in future studies. To my knowledge, this study, along with Alderson-Day and McGonigle-Chalmers' (2011) study, are the only studies to adapt the 20-Questions task into a game-like structure using a hinged board with 24 novel characters. Where this present study makes a unique contribution, is the task was created with the aim of assessing whether it had potential for normative performance characteristics (including scores and typical patterns of responding) to be established in future versions of this test, in order to capture normal variation of concept formation in children. A sample of 'typically developing' children were recruited into this study to allow for an exploration of these children's behaviours and strategies during the task. Additionally, children of different ages were tested and compared to introduce a level of complexity to the data.

By using a multi-trial format, this test not only required the children to use their concept formation and problem solving abilities, but it also involved working memory. Although the children were able to physically eliminate irrelevant Aliens, they still had to remember which questions they had asked before, while making a hypothesis about the future. Thus, in using a multi-trial format, children were required to form concepts based on not only what they could see, but what has happened in the recent episodic past. Therefore, the Alien Game patently involves both concept formation and hypothesis formation, and this is what makes the multi-trial format more ecologically valid than single trial tests.

At this first phase of test development, this study identified that similar strategies were used by the typically developing children when playing the Alien Game. This demonstrated its usefulness in capturing 'norms'. It was also found that the children enjoyed the Alien Game more than the Matrix Reasoning task and significantly more than the Similarities tasks. It could be argued that the Matrix Reasoning task was justifiably more popular than Similarities because this task is

visual and only required a participant to point to their answer, rather than having to think of a way to verbally express their answer. However, the Alien Game was the most popular despite also depending upon verbal abilities.

The test materials were designed to be appropriate for the children's developmental level. Previous research has often used measures which are standardised for both adults and children, rather than being specifically designed to engage children. It was felt that a test designed specifically for children could enhance engagement and thus the interpretation of how children form concepts while problem solving. The fact that children enjoyed the Alien Game appeared to also help build rapport and all children were observed to maintain attention throughout the task. Additionally, all of the children could play the game, including the child who had never played a 20-Questions or similar game before. Within this study, the standardised instructions appeared enough for him to get a grasp of the game overall, however, a couple of his questions yielded no items to be eliminated, but he noticed his mistake on both occasions. Each child was asked "Do you have any questions?" before starting also and was given an opportunity to state if they were unsure of what to do, which none of them did.

The advantages of basing the number of Alien's characteristics on a game-like structure is that it both enhanced the task's game-like quality, and allowed this test for children potentially using 'riskier' strategies. If the game was set up so that every time a child asked a question it removed half of the items, it would have been potentially easier for the researcher to score, but there would have been more predictability to the game; chance is what gives this task a game-like component.

Furthermore this task was also designed to be accessible to colour-blind children. It was also designed to be to be culturally fair. However, this will need to be verified in future research (see section. 4.7.)

4.4.1.2. *Codes* - a strength of this study was the two-phase coding system used to generate the codes. By basing the first phase of codes on previous research, this increased the generalizability of the study. The second phase of codes were

generated from the data itself. The second phase was important as this study is a new test of concept formation and so codes from previous studies needed to be adapted, and new codes created, in order to best understand the strategies used by the children on this task. Some codes which were not applicable to the present dataset, such as 'Repetition immediate', were kept in the coding system as previous research has indicated that these behaviours might be more common among samples of children with concept formation difficulties (e.g. Alderson-Day & McGonigle-Chalmers, 2011; Baldo et al., 2004).

The new coding scheme designed for this game identified that when given a dataset whereby each question yielded a different number of items being removed, the majority of these children sometimes used a strategy which could be deemed as 'risky' at least once. This finding may not have been identified if the characteristics chosen could yield removal of half of the items with each question or if the codes only captured 'Constraint seeking', 'Hypothesis seeking', and 'Pseudo-constraint-seeking' questions. It is not clear why the children in this study asked a 'riskier' question when they did. Perhaps this was an error, or a result of a child not sufficiently surveying the scene, but it is also possible that they used this as an intentional strategy to attempt to remove a greater number of items at once. Whether or not these strategies were intentional, what was observed among this sample, was that the "riskier" questions were found among other, more conservative, strategies. This finding is unique to this study and its scoring system.

#### 4.4.2. Limitations of the Current Study

Despite the strengths outlined above, there were some limitations with the study. Firstly, as this exploratory study used an opportunity sample, there were a number of limitations with the sample selected. This includes the sample being small (n=18), meaning that further research with larger groups will need to be conducted (see section 4.7.). However, at this pilot, exploratory stage of test development, this study was designed to see whether there were any indicators that this test could be used as a neuropsychological test of concept formation before a more in depth analysis using a larger number of participants occurred. This sample also only had two children with English as an additional language,

which is not representative of the current population in London. Additionally, no 6 year old children were recruited as the sample was recruited from classes within schools where the researcher had personal contacts.

The original pictures of the aliens were shown to a group of adults prior to the study commencing. Perhaps pre-testing these with a group of children instead for correct identification or appropriate prominence of certain features could have been helpful. This may have also led to changes being made to the pictures, such as perhaps changing the size of some of the features that were asked about less frequently (e.g. tails).

The interpretation of the results was also complicated by this sample of children having above average scores on the WISC-IV subtests, indicating that this group of children performed above the level expected for their age on these tasks. The fact that these children obtained good results on the WISC-IV subtests could also be related to the recruitment methods. Different schools opted for different approaches to opt children into the study. However, whether the parents opted their child into the study, or whether the school chose children to be opted in, both methods may have led to children with particular characteristics being recruited. For example, the children opted in by their parents may be from families with certain types of characteristics or values or where the teacher acted in *Loco parentis*, the children may have been chosen by the teacher for being a 'bright' child who was trusted by the teacher to do well in the task. The latter idea is supported by the fact that the children who were selected by the teacher, the 7-8 year olds, had particularly high scores on the WISC-IV subtests. The children in different age groups were also from different schools, in different areas of London, meaning that there could be additional confounding variables between the children in different age groups. There was also a different number of children in each age group (7-8 = n=4, 9-10 = 7, 11=7).

Although the game-like procedure is seen to have improved engagement, a negative of the game-like structure is that it introduced an element of chance, meaning that the results may not be exclusively based on a child's skill. While playing a game such as 'Guess Who?', a child can play against their older sibling

and still win sometimes due to the element of chance. Within the Alien Game there is a chance that someone could get lucky with their questions, for example, one participant guessed the correct colour of the Alien in their first question on each trial by chance. However, there are two aspects to this, as by asking about the alien's colour, the participant was asking a 'Constraint seeking' question, which has been identified as the best strategy to use, meaning that he already had a one in three chance of getting the colour right each time. Therefore, although it was by chance that he guessed the correct colour right first each time, he also used a 'good strategy' to enhance his chance of being in that position. The researcher aimed to reduce the impact of 'lucky guesses' on the child's score by using a scoring system which accounted for 'chance' and by having the same four target aliens in the same order each time for each child to keep the test consistent. Additionally, the minimum potential number of items removed by questions was used in the AS and IAS rather than the maximum to test for question quality.

As the children knew that the Alien Game was created by the researcher, this meant that the question asked by the Likert scale, which tested how enjoyable this task was, is slightly different for the Alien Game to the Likert scales for the Similarities or Matrix Reasoning tasks. However, if the children's answers were subject to task demand, we may have also expected to see a significant difference between the Matrix Reasoning task and the Alien Game.

It was perhaps unsurprising that the vast majority of questions were 'Constraint seeking' since participants were not able to eliminate just one item for the majority of the task. However, as many of the codes used in this study were not mutually exclusive, the fact that this code was given so often did not obscure other patterns of behaviour being observed and coded. Additionally, not allowing children to ask about one item at a time throughout the task can be seen as a strength of this study, as it encouraged children to form hypotheses.

A difficulty in general with 20-Questions style tasks is that they rely heavily on language and communication (Alderson-Day & McGonigle-Chalmers, 2011). Participants are required to generate verbal labels for categories in order to ask

the researcher questions to solve the problem. All the children in this study asked full and complete questions, which may not have happened in a non-clinical sample. The language demands of this study are difficult to reduce, given the nature of this task, and therefore, a requirement for using this task is that a child is able to communicate their answers. Within the present study, verbal abilities were tested using the Similarities task, and no child was measured as having difficulties in this area.

One of the children within the sample was unfamiliar with 20-Questions style games. Therefore, if this study were to be rerun, perhaps the children need to not only be asked if they have played a similar game before, but children should be offered a practice trial.

Compared with other standard tests of 20-Questions, which use a natural hierarchy of super-ordinate categories; this novel task was only able to offer groupings based on perceptual information (e.g. colour, shape). However, the strength here is that this aimed to make it a culturally fair test.

A further limitation is the extent to which clear inferences can be drawn from this task. As all of the children passed the task, and no Learning Slope was found, this could potentially indicate that the task was too 'easy' and the children's scores may have reached a 'ceiling affect'. However, it could also indicate that typically developing children can complete this task, whereas perhaps children with difficulties with concept formation may struggle more.

Despite the limitations discussed, the present study provides important new information regarding how typically developing children strategically form and use categories and concept formation to support their cognitive processes. It also found that typically developing children used similar patterns while playing this new test of concept formation, which indicates potential for normative characteristics to be established.

#### 4.4.3. Research Reflexivity

Being reflexive within research involves considering the impact of the circular relationship between the researcher, their history and social context, and the resulting research findings (Flanagan, 1981). During and after conducting this study, I reflected on the impact of my own social context and how it may have influenced my chosen methodology and interpretations of the results.

In the Methodology, I described adopting a 'critical realist' approach within this study. This approach is in line with my own curious position; not wanting to take anything for granted without questioning it. I feel that this stance has been shaped by both my family, who encouraged me to question things rather than take them at face value, as well as being influenced over the last three years by the thought-provoking and critical doctoral training that I am currently undertaking. In taking this approach with this particular topic, I acknowledge positioning the cognitive domains of executive functioning and concept formation as social constructs. However, alongside this view, I also attempted to measure the behavioural responses observed in response to testing these socially constructed categories. Within the current context of the NHS structure, access to resources are often given as a result of diagnosis, a trend also found within schools (e.g. in order for a child to get a statement of special educational needs/an education and health care plan). Thus, I acknowledged that in order for children to get appropriate support, we need to be able to measure a child's deficits and difficulties.

To me, creating a neuropsychological assessment was an obvious choice when aiming to evaluate and understand this construct, as this fits within the context of my role as a trainee clinical psychologist and my previous work as an assistant psychologist and research assistant. In these roles, I have used many neuropsychological tests in order to better understand a person's strengths, difficulties, and potential. However, my more 'critical' self questioned this method due to concerns regarding my contribution to an area of research, and thus social discourses, which reify 'impairments' in cognitive domains as being positioned as being fixed, internalised aspects of a person, rather than being viewed as context-dependent. Additionally, I have concerns that my test may support the

circular, self-perpetuating relationship with diagnoses which are frequently driven by economics rather than the individual and their family's best interests.

Conversely, the 'realist' in me balanced any apprehensions with my belief that concept formation needs to be better understood so that any difficulties can be captured early and to allow appropriate support and/or interventions to be put in place.

While reflecting on the methodology I chose, I acknowledge that this is just one way of investigating this construct. While this methodology arguably fits with the dominant approach taken in research in this area and within the context of the current scientific paradigm, this does not mean that this approach is the only route to testing concept formation. For example, if I had instead adopted a phenomenological perspective, I might have alternatively attempted to investigate the subjective experience of participants with difficulties in concept formation through using a qualitative approach to analysis, which might have resulted in different findings. In other words, I do not think there is a 'right' way to approach this research which would have resulted in an absolute truth or truly 'valid' result. Rather there are multiple approaches, perspectives, and interpretations from research which result from the multiple different contexts we exist within. I feel it is important to bear this in mind alongside the results and interpretations presented within this thesis.

#### **4.5. Conclusions**

The present study aimed to create a game-like structure, The Alien Game, that could be used as a test of concept formation, designed specifically for children. A two-phase content analysis was conducted and the results indicated that the typically developing children in this sample used similar strategies when completing this task. The results had some similarities with those from previous tests of concept formation, such as children predominantly used questions which aimed to divide the items into two large groups. This study also found that many children used strategies which could be seen as 'risky' at least once. The Alien

Game was scored using an AS, IAS, and WACH which gave global weighted scores.

The LS found that the children's scores stayed fairly consistent throughout the four trials and thus no significant difference was found. Additionally, the Time to First Question measure did not relate to the child's AS, IAS or WACH.

Overall, the Alien Game was rated as being an enjoyable test which engaged the children, whilst capturing typical strategies used by the children.

#### **4.6. Clinical Implications**

The finding that typically-developing children in this study did use similar strategies while being tested on the Alien Game indicates that normative characteristics can be established from using this test and it is thus worth continuing to develop this task into a neuropsychological test. The development of a neuropsychological tool of concept formation will be of benefit to children, parents, and those working with a child with concept formation difficulties in both clinical practice and future research. The availability of this data and findings of the present study will help take steps towards overcoming the current issue of there not being a test of concept formation designed specifically for children.

Children may have academic and social difficulties for a number of different reasons. Thus, being able to identify what difficulties someone has can help when putting in place appropriate interventions for a child. The earlier these difficulties can be identified, the earlier the interventions can be put in place.

Additionally, as this task was 'enjoyable', this could help to build rapport. The high enjoyment and engagement observed in this present study, along with the fact that there is relatively high success on this task, indicates that this test would be useful in engaging children who are sometimes harder to engage, such as those with head injuries. Within this population, children are likely to have an intersection of difficulties including social cognition and inattention difficulties (Donders & Kuldanek, 1998). This test accounts for this by being engaging and

using a game-like format. From my own experience of working in a specialist Attention Deficit Hyperactivity Disorder (ADHD) Child and Adolescent Mental Health Service, I observed how these children often struggled to maintain attention when being assessed using instruments such as the WISC-IV. The Alien Game is not only engaging, but it can be administered within a relatively short time scale meaning that it is less time-consuming for children. These aspects would make this test accessible to children with a range of dysexecutive difficulties, including children with foetal alcohol syndrome, ASD, and Edward's syndrome. However, in its current format, it would require children to have verbal abilities.

The simplicity of the test means that it is not just accessible to the test takers, but also to the test users. Within this present study, the second coder was not a psychologist, and had no problems in coding this data through using the information provided to him. This demonstrates that this test could be used by not just psychologists, but also by teachers within schools if need be. This test could also be used in various settings, including: a clinic, in schools, in a child's home, or at a child's bedside in rehabilitation settings.

The above information demonstrates the Alien Game's utility in being used as neuropsychological test which could help in the diagnosis of concept formation difficulties in children. Being able to identify concept formation difficulties at an early stage could lead to interventions being introduced earlier in a child's life.

#### **4.7. Future Research Directions**

This study was the initial stage of test development, which found that typically developing children used similar strategies when playing the Alien Game. Further research can now be carried out to continue the test development and assess the scope and consistency of strategies identified in the children's responses. This sub-section will discuss the future research directions which would support the development of norms, and establish the reliability and validity of this task, to ultimately work towards creating a standardised version of the Alien Game.

#### 4.7.1. Reliability

The reliability of the Alien Game could be assessed firstly through testing the internal, inter-item consistency through using a Cronbach's alpha to test whether the four trials were testing the same aspects of concept formation and problem solving. It would also be useful to conduct a test-retest study with the same group of children to see whether these children scored similarly both times. The researcher would conduct this study while being aware there is potential that there may be an element of practice, and thus the children's scores may increase. In order to account for this, the children could be given a parallel form of the task when they are re-tested, such as using pictures of 'Alien Vehicles' or another abstract design, which uses the same amount of characteristics as the original Alien Game. This would not only be useful in testing the reliability of the Alien Game, but could also be used in future within clinical settings, such as in rehabilitation services to assess children's progress.

This test could then be administered to a larger group of participants, which should include children aged 6 and younger. With the younger children, a different single trial test, appropriate for their age group would need to be used as a comparison to test whether any children have language or visuo-spatial difficulties, such as Picture Concepts in the Wechsler Preschool and Primary Scale of Intelligence. These studies should be conducted in a way that reduces possible confounding variables within a sample, such as testing children in a similar location, for example, if possible, testing everyone who meets the inclusion criteria in one class in each year group of a school, or randomly selecting a number of children from each school. Also, within a larger group, it would be useful to see whether the results from this task correlate more with the Similarities, Matrix Reasoning, or with other multi-trial test of concept formation.

In order to test whether the children's Likert scale responses were impacted by task demand within this study (as the children knew that the researcher had created the game), a future study could be conducted by someone other than the present researcher to test whether the children's enjoyment scores differed.

#### 4.7.2. Validity

In order to test for validity, future research could test the concurrent validity of the study by testing the children's results from the Alien Game with their results from other measures. Within the present study, the Similarities and Matrix Reasoning WISC-IV subtests were used to assess verbal and visuo-spatial concept formation deficits. These subtests were chosen as the WISC-IV is commonly used and both tests are considered to be single trial tests of concept formation, so no practice effects were expected to occur between these tasks and the multi-trial Alien Game. It would be useful to compare a child's performance on the Alien Game with another multi-trial measure of concept formation. A difficulty in choosing an appropriate task for comparison is that the researcher is unaware of another standardised multi-trial task designed specifically for children. Therefore, in order to test for convergent validity, the Alien Game would need to be compared with other multi-trial tasks designed for adults as well as children, such as the WCST. It would be useful to see which, if any, of the scores (AS, IAS, or WAch) correlated with other measures of concept formation. A further recommendation for a future study is to see if there are differences that emerge from comparing this test to a 20-Questions task that used more 'even' structure (where each question could possibly remove half of the items).

Future research may also compare the coding system used in this study, in which codes were not mutually exclusive from each other, with a test with a coding scheme whereby one code takes precedence to see which framework best captures the children's problem solving strategies. This is because while qualitative analysis allows for codes that are not mutually exclusive (Tesch, 1990), if a mutually exclusive coding system is found useful, a future study could be adapted to use it.

Research should also recruit children from a 'known group' sample, using groups of children who were known to have executive function problems, for example, children with frontal lobe injuries after a road traffic accident or another ABI, ASD, and ADHD. This is similar to the ecological validity testing of the Brixton Spatial Anticipation Test conducted between people with frontal lesions and controls (Burgess & Shallice, 1997). It is also possible that, from these new studies, new

codes may need to be developed to more accurately capture these children's behaviour during the Alien Game. Future research should also test for criterion validity by correlating performance on the Alien Game with some objective measure which captures a child's real world performance, such as a measure of executive functioning which captures successful performance in school, including abilities in planning, organisation, and self-control.

This test was created with the aim of being a culturally fair measure. It would thus be important to test this by conducting this study with children from different cultures and who speak different languages. It would also be useful to find out whether people from other cultures are familiar with 20-Questions style games, and whether this affects their performance on this task if they do not.

After this future research has been conducted it would be useful to revisit whether a cut-off point of 10 is the most suitable pass mark.

#### **4.8. Concluding Summary**

This thesis has explored the utility of using a game-like procedure in a task specifically aimed at testing children's concept formation abilities. As typical patterns of responding were found among the typically developing children in this study, these findings can now be used as the basis for further research with the aim of further developing the Alien Game into a useful tool to help the early identification of difficulties in children's concept formation and problem solving abilities.

## REFERENCES

- Ahn, W. K., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 391. <https://doi.org/10.1037/0278-7393.18.2.391>
- Albert, D., Opwis, K., & Regard, M. (2010). Effect of drawing hand and age on figural fluency: A graphomotor study with the five-point test in children. *Child Neuropsychology*, *16*(1), 32-41. <http://dx.doi.org/10.1080/09297040903049061>
- Albert, M. L., & Knoefel, J. E. (2011). *Clinical neurology of aging* (3<sup>rd</sup> ed.). New York: Oxford University Press.
- Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is it a bird? Is it a plane? Category use in problem-solving. *Journal of Autism and Developmental Disorders*, *41*, 555–565. <https://doi.org/10.1007/s10803-010-1077-9>
- Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological Rehabilitation*, *8*(3), 319–349. <https://doi.org/10.1080/713755568>
- Armstrong, D. M. (1973). *Belief, truth and knowledge*. Cambridge: Cambridge University Press.
- Baldo, J. V., Delis, D. C., Wilkins, D. P., & 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. [https://doi.org/10.1016/S0887-6177\(03\)00074-X](https://doi.org/10.1016/S0887-6177(03)00074-X)
- Barceló, F. (2001). Does the Wisconsin card sorting test measure prefrontal function? *The Spanish Journal of Psychology*, *4*(1), 79-100. <https://doi.org/10.1017/S1138741600005680>
- Barrett, S. E., Abdi, H., Murphy, G. L., & Gallagher, J. M. (1993). Theory-based correlations and their role in children's concepts. *Child Development*, *64*, 1595-1616. <https://doi.org/10.2307/1131458>
- Bekman, N. M., Goldman, M. S., Worley, M. J., & Anderson, K. G. (2011). Pre-adolescent alcohol expectancies: Critical shifts and associated

- maturational processes. *Experimental and clinical psychopharmacology*, 19(6):420–432. <https://doi.org/10.1037/a0025373>
- Bernstein, J. H. (2000). Developmental neuropsychological assessment. In K. O. Yeates, M. D. Ris, & H. G. Taylor (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (pp. 405-438). New York: Guilford.
- Bhaskar, R.A. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy*. London: Verso.
- Bowden, S. C., Fowler, K. S., Bell, R. C., Whelan, G., Clifford, C. C., Ritter, A. J., & Long, C. M. (1998). The reliability and internal validity of the Wisconsin card sorting test. *Neuropsychological Rehabilitation*, 8, 243-254. <https://doi.org/10.1080/713755573>
- Braun, J. S. (1963). Relation between concept formation ability and reading achievement at three developmental levels. *Child Development*, 34, 675-668. <https://doi.org/10.2307/1126760>
- Brown, R. W. (1956) Language and categories. In J. J. Bruner, J. S. Jerome, J. J. Goodnow, & G. A. Austin (Eds.), *A study of thinking* (Appendix). New York: Wiley.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. Oxford, England: Wiley
- Burgess, P. W., & Shallice, T. (1997). *Brixton spatial anticipation test*. Suffolk, UK: Thames Valley Test Co Ltd.
- Bynner, J. M. (1997). Basic skills in adolescents' occupational preparation. *The Career Development Quarterly*, 45, 305–321. <https://doi.org/10.1002/j.2161-0045.1997.tb00536.x>
- Callanan, M., & Oakes, L. M. (1989). Parent-child collaboration in the development of category inferences and causal theories: *Paper presented at the biennial meeting of the Society for Research in Child Development*, Kansas City, MO.
- Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F., & Segall, M. J. (2008). Evidence-based assessment of cognitive functioning in pediatric psychology. *Journal of Pediatric Psychology*, 33(9): 999–1014. <https://doi.org/10.1093/jpepsy/jsm138>
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

- Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences*, 3(3), 106-116. <https://doi.org/10.3844/jssp.2007.106.116>
- Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology, Vol. 1: Body and mind* (pp. 706-743). New York: Oxford University Press.
- Carmichael, C. A., & Hayes, B. K. (2001). Prior knowledge and exemplar encoding in children's concept acquisition, *Child Development*, 72(4), 1071-1090. <https://doi.org/10.1111/1467-8624.00335>
- Castellanos, I., Kronenberger, W. G., Beer, J., Colson, B. G., Henning, S. C., Ditmars, A., & Pisoni, D. B. (2015). Concept formation skills in long-term cochlear implant users. *Journal of Deaf Studies and Deaf Education*, 20(1), 27-40. <https://doi.org/10.1093/deafed/enu039>
- Cattell, R. B. & Cattell, H. E. P. (1973). *Measuring intelligence with the culture fair tests*. Champaign, IL: Institute for Personality and Ability Testing.
- Channon, S., & Crawford, S. (1999). Problem-solving in real-life-type situations: The effects of anterior and posterior lesions on performance. *Neuropsychologia*, 37, 757–770. [https://doi.org/10.1016/S0028-3932\(98\)00138-9](https://doi.org/10.1016/S0028-3932(98)00138-9)
- Charlesworth, R., & Lind, K. K. (2010). *Math and science for young children* (6th ed.). Albany: Delmar Cengage Learning.
- Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin card sorting test. *Journal of Clinical and Experimental Neuropsychology*, 8, 219-228. <https://doi.org/10.1080/01688638608401314>
- Chelune, G. J., & Thompson, L. L. (1987). Evaluation of the general sensitivity of the Wisconsin card sorting test among younger and older children. *Developmental Neuropsychology*, 3, 81-89. <https://doi.org/10.1080/87565648709540365>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Darby, D. & Walsh, K. W. (2005). *Walsh's neuropsychology: A clinical approach*. Edinburgh: Elsevier Churchill Livingstone.

- Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis–Kaplan Executive Function System [D-KEFS]*. San Antonio: The Psychological Corporation, Harcourt Brace & Company.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan executive function system: An update. *Journal of the International Neuropsychological Society*, *10*(2), 301-303.  
<https://doi.org/10.1017/S1355617704102191>
- Dennis, M. (1988). Language and the young damaged brain. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 85-123). Washington, DC: American Psychological Association.
- Donders, J., & Kuldane, A. (1998). Traumatic brain injury. In R. Ammerman & J. Campo (Eds.), *Handbook of pediatric psychology and psychiatry* (pp. 166-190). Needham Heights, MA: Allyn & Bacon.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., . . . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446.  
<https://10.1037/0012-1649.43.6.1428>
- Edwards, O. W., & Oakland, T. D. (2006). Factorial invariance of Woodcock-Johnson III scores for African Americans and Caucasian Americans. *Journal of Psychoeducational Assessment*, *24*(4), 358-366.  
<https://doi.org/10.1177/0734282906289595>
- Elliott, R. (2003). Executive functions and their disorders. *British Medical Bulletin*, *65*(1), 49–59. <https://doi.org/10.1093/bmb/65.1.49>
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science*, *146*, 668–670.
- Farrar, M. J., Raney, G. E., & Boyer, M. E. (1992). Knowledge, concepts, and inferences in childhood. *Child Development*, *63*(3) 673–691.  
<https://doi.org/10.2307/1131354>
- Ferrier, J. F. (1854). *Institutes of metaphysics: The theory of knowing and being*. Edinburgh: Blackwood.
- Flanagan, O. J. (1981). Psychology, progress, and the problem of reflexivity: A study in the epistemological foundations of psychology. *Journal of the History of the Behavioural Sciences*, *17*(3), 375-386.

[https://doi.org/10.1002/1520-6696\(198107\)17:3<375::AID-JHBS2300170308>3.0.CO;2-U](https://doi.org/10.1002/1520-6696(198107)17:3<375::AID-JHBS2300170308>3.0.CO;2-U)

- Flavell, J. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.
- Floyd, R. G., McCormack, A. C., Ingram, E. L., Davis, A. E., Bergeron, R., & Hamilton, G. (2006). Relations between the Woodcock-Johnson III clinical clusters and measures of executive functions from the Delis-Kaplan executive function system. *Journal of Psychoeducational Assessment, 24*, 303-317. <https://doi.org/10.1177/0734282906287823>
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171-191. <https://doi.org/10.1037/0033-2909.101.2.171>
- Friedman, N., Miyake, A., Young, S., DeFries, J., Corley, R., & Hewitt, J. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General 137*(2), 201–225. <https://doi.org/10.1037/0096-3445.137.2.201>
- Fodor, J. (1981). The present status of the innateness controversy. In J. Fodor (Ed.), *Re-presentations*, (pp. 257-316). Cambridge, MA: MIT Press.
- Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.
- Gardiner, R. (1958). Alfred A. Strauss, 1897-1957, *Exceptional Children, 24*, 373-374.
- Gelman, R. (1978). Cognitive development. *Annual Review of Psychology, 29*, 297-332. <https://doi.org/10.1146/annurev.ps.29.020178.001501>
- Gelman, R. (2002). *Cognitive development*. In H. Pashler & D. L. Medin (Eds.), *Stevens' handbook of experimental psychology* (3<sup>rd</sup> ed., pp. 575-630). Hoboken, NJ: Wiley.
- Gelman, R., & Au, T. K. F. (1996). *Handbook of perceptual and cognitive development*. San Diego, CA: Academic Press.
- Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell, & E. M. Markman (Eds.), *Handbook of child psychology vol. III: Cognitive development* (4<sup>th</sup> ed., pp. 167-230). New York: Wiley.
- Gelman, S. A. (Ed.). (1999). Concept development in preschool children. *Paper presented at the Forum on Early Childhood Science, Mathematics, and Technology Education*, Washington, DC. *Dialogue on early childhood*

- science, mathematics, and technology education* (pp. 50-61). Washington, DC: American Association for the Advancement of Science/Project 2061.
- Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60*, 115–140. <https://doi.org/10.1146/annurev.psych.59.103006.093659>
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-208. [https://doi.org/10.1016/0010-0277\(86\)90034-X](https://doi.org/10.1016/0010-0277(86)90034-X)
- Gioia, G. A. (2015) Multimodal evaluation and management of children with concussion: Using our heads and available evidence. *Brain Injury, 29*(2), 195-206. <https://doi.org/10.3109/02699052.2014.965210>
- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology, 6*(3), 235-238. <https://doi.org/10.1076/chin.6.3.235.3152>
- Gligorović, M., & Buha, N. (2013). Conceptual abilities of children with mild intellectual disability: Analysis of Wisconsin card sorting test performance, *Journal of Intellectual & Developmental Disability, 38*(2), 134–140. <https://doi.org/10.3109/13668250.2013.772956>
- Glutting, J. J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test manual*, Wilmington, DE: Wide Range.
- Goldberg, S., Perlmutter, M., & Myers, N. (1974). Recall of related and unrelated lists by 2 year olds. *Journal of Experimental Child Psychology, 18*, 1-8. [https://doi.org/10.1016/0022-0965\(74\)90083-6](https://doi.org/10.1016/0022-0965(74)90083-6)
- Goldstein, K., & Scheerer, M. (1941). Abstract and concrete behavior: An experimental study with special tests. *Psychology Monograph, 52*(2), 1-151. <https://doi.org/10.1037/h0093487>
- Gorham, D. R. (1956). A proverbs test for clinical and experimental use. *Psychological Reports, 2*(1), 1-12. <https://doi.org/10.2466/PR0.2..1-12>
- Greve, K. W., Farrell, J. F., Besson, P. S., & Crouch, J. A. (1995). A psychometric analysis of the California Card Sorting Test. *Archives of Clinical Neuropsychology, 10*(3), 265-78. [https://doi.org/10.1016/0887-6177\(94\)00046-S](https://doi.org/10.1016/0887-6177(94)00046-S)
- Guthrie, J. T., Wigfield, A., Barbosa, P. Perencevich, K. C. Taboada, A., Davis, M. H., Scafiddi, N. T., & Tonks, S. (2004). Increasing reading

- comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, 96(3), 403-423.  
<https://doi.org/10.1037/0022-0663.96.3.403>
- Harper, D. (2011). Choosing a qualitative research method. In D. Harper & A. R. Thompson (Eds.), *Qualitative research methods in mental health and psychotherapy* (pp. 83-98). West Sussex, United Kingdom: Wiley-Blackwell.
- Harris, P.L., & Koenig, M.A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, 77, 505–24.  
<https://doi.org/10.1111/j.1467-8624.2006.00886.x>
- Hashimoto, N., McGregor, K. K., & Graham, A. (2007). Conceptual organization at 6 and 8 years of age: Evidence from the semantic priming of object decisions. *Journal of Speech, Language, and Hearing Research*, 50(1), 161-176. [https://doi.org/10.1044/1092-4388\(2007/014\)](https://doi.org/10.1044/1092-4388(2007/014))
- Headway, The Brain Injury Association (2017). Executive dysfunction after brain injury. Retrieved from: <https://www.headway.org.uk/about-brain-injury/individuals/effects-of-brain-injury/executive-dysfunction/>.
- Heaton, R., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test (WCST) manual revised and expanded*. Odessa, FL: Psychological Assessment Resources.  
<https://10.1037/t31298-000>
- Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1264-1282. <https://doi.org/10.1037/0278-7393.20.6.1264>
- Hill, R. W., Huelsman, T. J., Furr, R. M., Kibler, J., Vicente, B. B., & Kennedy, C. (2004). A new measure of perfectionism: The perfectionism inventory. *Journal of personality assessment*, 82(1), 80-91.  
[https://doi.org/10.1207/s15327752jpa8201\\_13](https://doi.org/10.1207/s15327752jpa8201_13)
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks preferential attachment or preferential acquisition?. *Psychological Science*, 20(6), 729-739.  
<https://doi.org/10.1111/j.1467-9280.2009.02365.x>
- Huizinga, M., & van der Molen, M. W. (2007). Age-group differences in set-switching and set-maintenance on the Wisconsin card sorting task.

- Developmental Neuropsychology*, 31(2), 193-215.  
<https://doi.org/10.1080/87565640701190817>
- Hunt, E. (1962). *Concept learning: An information processing problem*. New York: John Wiley and Co.
- Inhelder, B., & Piaget, J. (1964). *Classification and seriation*. London: Routledge & Kegan Paul.
- Jacques, S., & Zelazo, P. D. (2001). The flexible item selection task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, 20, 573 – 591.  
<https://doi.org/10.1207/875656401753549807>
- Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development*, 1(1), 73-96. [https://doi.org/10.1016/S0885-2014\(86\)80024-7](https://doi.org/10.1016/S0885-2014(86)80024-7)
- Keil, F. C. (1989). *Concepts, kinds and development*. Cambridge, MA: Bradford Books/MIT Press.
- Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, 65, 103-135. [https://doi.org/10.1016/S0010-0277\(97\)00041-3](https://doi.org/10.1016/S0010-0277(97)00041-3)
- Kizilbash, A., & Donders, J. (1999). Latent structure of the Wisconsin card sorting task after pediatric traumatic brain injury. *Child Neuropsychology*, 5(4), 224-229. [https://doi.org/10.1076/0929-7049\(199912\)05:04;1-R;FT224](https://doi.org/10.1076/0929-7049(199912)05:04;1-R;FT224)
- Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, 13(1), 111-124.  
<https://doi.org/10.1017/S0954579401001080>
- Krascum, R. M., & Andrews, S. (1998). The effects of theories on children's acquisition of family-resemblance categories. *Child Development*, 69(2), 333-346. <https://doi.org/10.2307/1132168>
- Langer, S. (1967). *An introduction to symbolic logic*. New York: Dover.
- Levi, A. (1965). Treatment of a disorder of perception and concept formation in a case of school failure. *Journal of Consulting Psychology*, 29(4), 289-295.  
<https://doi.org/10.1037/h0022425>
- Lezak, M. D. (1995). *Neuropsychological assessment*. New York: Oxford University Press.

- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment*, (4th ed.). New York: Oxford University Press.
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*, 3-28.  
<https://doi.org/10.1037/0096-3445.130.1.3>
- Lopez, B., Lincoln, A., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders*, *35*(4), 445–460.  
<https://doi.org/10.1007/s10803-005-5035-x>
- Lord-Maes, J., & Obrzut, J. E. (1997). Neuropsychological consequences of traumatic brain injury in children and adolescents. In E. D. Bigler & E. Clark (Eds.), *Childhood traumatic brain injury: Diagnosis, assessment, and intervention* (pp. 101–114). Austin, TX: Pro-Ed.
- Lorsbach, T. C., Wilson, S., & Reimer, J. F. (1996). Memory for relevant and irrelevant information: Evidence for deficient inhibitory processes in language/learning disabled children. *Contemporary Educational Psychology*, *21*, 447–466. <https://doi.org/10.1006/ceps.1996.0030>
- Luria, A. R. (1966). *Higher Cortical Functions in Man*. New York, Basic Books.
- Macario, J. F., Shipley, E. F., & Billman, D. O. (1990). Induction from a single instance: Formation of a novel category. *Journal of Experimental Child Psychology*, *50*(2), 179-199. [https://doi.org/10.1016/0022-0965\(90\)90038-A](https://doi.org/10.1016/0022-0965(90)90038-A)
- Maddox, W. T., Filoteo, J. V. Glass, B. D., & Markman, A. B. (2010). Regulatory match effects on a modified Wisconsin card sort task. *Journal of the International Neuropsychological Society*, *16*, 352–359.  
<https://doi.org/10.1017/S1355617709991408>
- Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 250-269.  
[https://doi.org/10.1016/S0022-5371\(84\)90170-1](https://doi.org/10.1016/S0022-5371(84)90170-1)
- Mandler, J. M. (2004). *The foundations of mind*. New York: Oxford University Press.
- Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology*, *21*, 207–230. <https://doi.org/10.1080/09515080801980179>

- Markman, E. M. & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs thematic relations. *Cognitive Psychology*, 16, 1-27. [https://doi.org/10.1016/0010-0285\(84\)90002-1](https://doi.org/10.1016/0010-0285(84)90002-1)
- Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language & Communication Disorders*, 43(2), 181-200. <https://doi.org/10.1080/16066350701340719>
- McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, 3, 341-350. [https://doi.org/10.1016/0010-0277\(74\)90003-1](https://doi.org/10.1016/0010-0277(74)90003-1)
- McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McShane, J. (1980). *Learning to talk*. Cambridge, MA: Cambridge University Press.
- McShane, J., & Dockrell, J. (1983). Lexical and syntactic development. In B. Butterworth (Ed.), *Speech production Vol. II*, London: Academic Press.
- Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35, 113-138. <https://doi.org/10.1146/annurev.ps.35.020184.000553>
- Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorisation* (pp. 201–233). Cambridge: Cambridge University Press.
- Mesulam, M. M. (2000). *Principles of behavioral and cognitive neurology*. Oxford University Press.
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological psychology*, 26(1), 3-67. [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- Minshew, N., Meyer, J., & Goldstein, G. (2002). Abstract reasoning in autism: A dissociation between concept formation and concept identification. *Neuropsychology*, 16(3), 327–334. <https://doi.org/10.1037/0894-4105.16.3.327> .
- Minshew, N., Siegel, D., Goldstein, G., & Weldy, S. (1994). Verbal problem solving in high functioning autistic individuals. *Archives of Clinical*

- Neuropsychology*, 9(1), 31–40. [https://doi.org/10.1016/0887-6177\(94\)90012-4](https://doi.org/10.1016/0887-6177(94)90012-4)
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6(7). doi:10.1371/journal.pmed1000097
- Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsychologia*, 31, 1367–1378. [https://doi.org/10.1016/0028-3932\(93\)90104-8](https://doi.org/10.1016/0028-3932(93)90104-8)
- Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. Olver, & P. Greenfield, (Eds.), *Studies in cognitive growth* (pp. 86–102). New York: Wiley.
- Murphy, G. L. (1993). Theories and concept formation. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 173-200). New York: Academic Press.
- Nelson, K. (1973). Some evidence for the cognitive primacy of categorization and its functional basis. *Merrill--Palmer Quarterly*, 19, 21-39.
- Nenty, H. J. (1986). Cross-cultural bias analysis of Cattell culture fair intelligence test: 67<sup>th</sup> Annual Meeting of the American Educational Research Association. San Francisco, CA, April 16-20.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287-293. <https://doi.org/10.1111/j.1745-3984.1985.tb01065.x>
- Piaget, J. (1951). *Play, dreams and imitation in childhood*, New York: Norton.
- Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). New York: Basic Books. (Original work published 1927).
- Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of research in science teaching*, 2(3), 176-186.
- Piaget, J., & Szeminska, A. (1952). *The child's conception of number* (C. Gattegno & F. M. Hodgson, Trans.). New York: Humanities Press. (Original work published 1941).
- Quine, W. V. O. (1977). Natural kinds. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press.

- Rapaport, D. (1946). *Diagnostic psychological testing. Vol. I*. Chicago: Year Book Publishers.
- Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive*. (Unpublished master's thesis). University of London, London.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford, United Kingdom: Oxford Psychologists Press. <https://doi.org/10.1037/t07027-000>
- Reichard, S., Schneider, M., & Rapaport, D. (1943). The development of concept formation in children. *Transactions of the Kansas Academy of Science (1903)*, *46*, 220-223. <https://doi.org/10.1111/j.1939-0025.1944.tb04860.x>
- Remine, M. D., Care, E., & Brown, P. M. (2008). Language ability and verbal and nonverbal executive functioning in deaf students communicating in spoken English. *Journal of Deaf Studies and Deaf Education*, *13*(4), 531-545. <https://doi.org/10.1093/deafed/enn010>
- Rey, G. (1983). Concepts and stereotypes. *Cognition*, *15*, 237-262. [https://doi.org/10.1016/0010-0277\(83\)90044-6](https://doi.org/10.1016/0010-0277(83)90044-6)
- Riccuiti, H. N. (1965). Object grouping and selective ordering in infants 12 to 24 months. *Merrill-Palmer Quarterly*, *11*, 129-148.
- Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, *27*, 313–328. <https://doi.org/10.2307/145737>
- Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale, (Leiter-3)*. Los Angeles: Western Psychological Services.
- Ropar, D., & Peebles, D. (2007). Sorting preference in children with autism: The dominance of concrete features. *Journal of Autism and Developmental Disorders*, *37*(2), 270–280. <https://doi.org/10.1007/s10803-006-0166-2>
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates
- Rose, S. A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. *Child Development*, 499-502. <https://doi.org/10.2307/1127977>

- Rossi, E. L., & Rossi, S. I. (1965). Concept utilization, serial order, and recall in nursery school. *Child Development, 36*, 771-779.  
<https://doi.org/10.2307/1126922>
- Satz, P. & Mogel, S. (1962). An abbreviation of the WAIS in clinical use. *Journal of Clinical Psychology, 18*, 77-79. [https://doi.org/10.1002/1097-4679\(196307\)19:3<298::AID-JCLP2270190311>3.0.CO;2-G](https://doi.org/10.1002/1097-4679(196307)19:3<298::AID-JCLP2270190311>3.0.CO;2-G)
- Scott, M. S., Greenfield, D. B., & Urbano, R. C. (1985). A comparison of complementary and taxonomic utilization: Significance of the dependent measure. *International Journal of Behavioral Development, 8*, 241–256.  
<https://doi.org/10.1177/016502548500800301>
- Shu, C., Lung, W., Tien, A., & Chen, C. (2001). Executive function deficits in non-retarded autistic children. *Autism, 5*(2), 165–174.  
<https://doi.org/10.1177/1362361301005002006>
- Siegler, R. S. (1977). The twenty questions game as a form of problem solving. *Child Development, 48*, 395–403. <https://doi.org/10.2307/1128632>
- Smidts, D. P., Jacobs, R., & Anderson, V. (2004). The object classification task for children (OCTC): A measure of concept generation and mental flexibility in early childhood. *Developmental Neuropsychology, 26*, 385–401. [https://doi.org/10.1207/s15326942dn2601\\_2](https://doi.org/10.1207/s15326942dn2601_2)
- Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology, 28*(2), 249-257.  
[https://doi.org/10.1016/0022-0965\(79\)90087-0](https://doi.org/10.1016/0022-0965(79)90087-0)
- Smith, L. B., Jones S. S., & Landau B. (1996). Naming in young children: A dumb attentional mechanism?. *Cognition, 60*, 143–71.  
[https://doi.org/10.1016/0010-0277\(96\)00709-3](https://doi.org/10.1016/0010-0277(96)00709-3)
- Starkey, D. (1979). The origins of concept formation: Object sorting and object preference in early infancy. *Child Development, 52*, 489-497.  
<https://doi.org/10.2307/1129166>
- Sweetland, R. C., & Keyser, D. J. (Eds.). (1991). *A comprehensive reference for assessment in psychology, education, and business* (3rd ed.). Austin, TX: Proed.
- Taylor, H. G., & Fletcher, J. M. (1990). Neuropsychological assessment of

- children. In M. Hersen & G. Goldstein (Eds.), *Handbook of psychological assessment* (2<sup>nd</sup> ed., pp 228-255). New York: Plenum.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Bristol, PA: Falmer Press.
- Trochim, W. (2000). *The research methods knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.
- Urbina, S. (2011). Tests of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge Handbook of Intelligence* (pp. 20-38). New York: Cambridge University Press.
- van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, 35, 427 – 449. <https://doi.org/10.1016/j.jecp.2003.12.002>
- Van Lancker, D. (1990). The neurology of proverbs. *Behavioural Neurology*, 3(3), 169-187. <https://doi.org/10.1155/1990/261504>
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic, *Family Medicine*, 37(5), 360-363.
- Vygotsky, L. S. (1986). *Thought and language*. (A. Kozulin Trans.), Cambridge, MA.: MIT Press. (Original work published 1934).
- Wanner, E., & Gleitman, L. (1983). *Language acquisition: The state of the art*. Cambridge, England: Cambridge University Press.
- Weber, R.P. (1990). *Basic Content Analysis* (2<sup>nd</sup> Ed.). Newbury Park, CA: Sage Publications.
- Wechsler, D. (2003). *The Wechsler intelligence scale for children [WISC-IV]* (4<sup>th</sup> ed.). San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale [WISC-IV]* (4<sup>th</sup> ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2014). *The Wechsler intelligence scale for children [WISC-V]* (5<sup>th</sup> ed.). San Antonio, TX: The Psychological Corporation.
- Wechsler D., Kaplan E., Fein D., Kramer J., Morris R., Delis D., Maerlender A. (2004). *Wechsler Intelligence Scale for Children (4<sup>th</sup> ed.)*. - *Integrated*. San Antonio, TX: Harcourt Assessment.

- Weigl, E. (1941). On the psychology of so-called processes of abstraction. *Journal of Abnormal Psychology and Social Psychology*, 36(1), 3-33. <https://doi.org/10.1037/h0055544>
- Weiss, A. A. (1964). The Weigl-Goldstein-Scheerer color-form sorting test: Classification of performance. *Journal of Clinical Psychology*, 20(1), 103-107. [https://doi.org/10.1207/s15516709cog1802\\_2](https://doi.org/10.1207/s15516709cog1802_2)
- Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In D. L. Molfese & V. J. Molfese (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139–186). Mahwah, NJ: Erlbaum.
- 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(2), 131-149. <https://doi.org/10.1080/87565649109540483>
- Williams, D., Goldstein, G., & Minshew, N. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology*, 12(4–5), 279–298. <https://doi.org/10.1080/09297040600681190>
- Willig, C. (2012). Perspectives on the epistemological bases for qualitative research. In H. Cooper (Ed.). *The handbook of research methods in psychology* (pp. 5-21). Washington: American Psychological Association.
- Willig, C. (2013). *Introducing qualitative research methods in psychology* (3rd ed.). Berkshire: Open University Press.
- Wing, L. (1996). *The autistic spectrum*. London: Constable.
- Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, 18(2), 221-281. [https://doi.org/10.1016/0364-0213\(94\)90002-7](https://doi.org/10.1016/0364-0213(94)90002-7)
- Wood, K. C., Smith, H., & Grossniklaus, D. (2001). Piaget's stages of cognitive development. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology*. Retrieved 30<sup>th</sup> March 2017, from: <http://projects.coe.uga.edu/epltt/> .
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside.

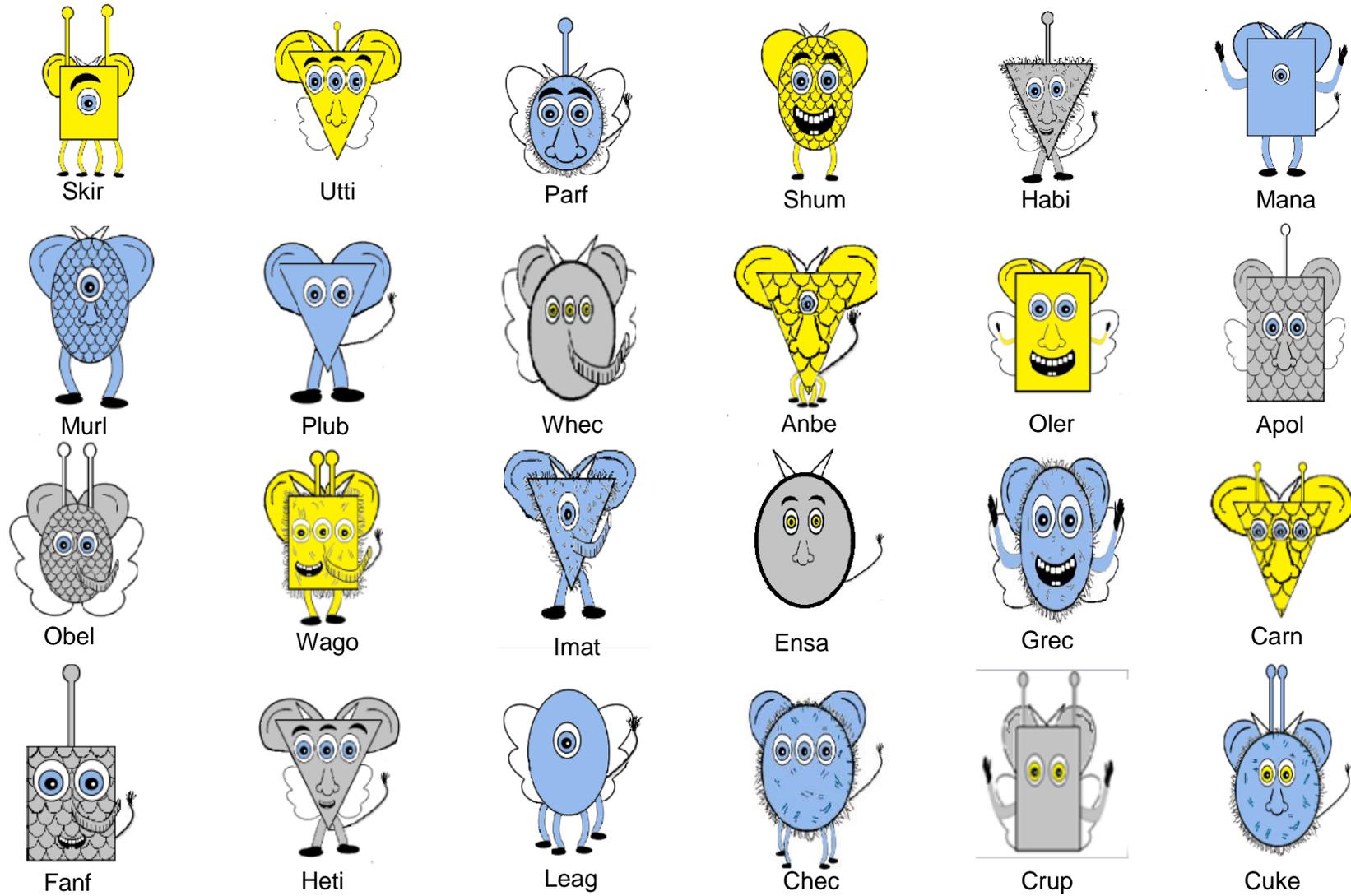
Yeates, K. O. & Taylor, H. G. (1998). Neuropsychological assessment of older children. In G. Goldstein, A. Puente, & E. Bigler (Eds.), *Human brain function, assessment and rehabilitation: Vol. III, Neuropsychology* (pp. 35-61). New York: Plenum.

Yeates, K. O, & Taylor, H. G. (2001). Neuropsychological assessment of children. In J. J. W. Andrews, D. H. Saklofske, & H. L. Janzen (Eds.), *Handbook of psychoeducational assessment: Ability, achievement, and behavior in children* (pp. 415-450). New York: Academic Press.

**APPENDICIES:**

- A. Pictures Used in the Alien Game.
- B. Consent Form – Child.
- C. Information Sheet.
- D. Opt Out Form For Parents/Guardians.
- E. Consent Form – Parents/Guardians.
- F. Head Teacher's *Loco Parentis* Form.
- G. Standardised Instructions For Child.
- H. Measure Of Task Acceptability Using Likert Scale.
- I. Coding Sheet.
- J. Coding Scheme Explained.
- K. Scoring For Code 7 and Code 8.
- L. Copy Of Ethics Application.
- M. Mean Learning Slope Total and in Different Age Groups.

**APPENDIX A:** Pictures Used in the Alien Game.



*Figure 3. Pictures of Aliens Created for the Alien Game*

**APPENDIX B: Consent Form – Child.**

**Professional Doctorate in Clinical Psychology.**



**Professional Doctorate in  
Clinical Psychology.**

**Consent Form – Child**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Researcher:** Annie Pavitt

I am studying how people solve problems while playing a game, and would like you to help me.

I will ask you to play a game that is similar to ‘Guess Who?’, except this will involve pictures of aliens. I will have already chosen an alien, and I would like you to ask yes/no questions in order to guess as quickly as you can which alien I have chosen.

After this, I will ask you to try two more tasks; one will involve solving puzzles, and the other will be answering a few questions about how two items are similar.

I will record what you say in the game and will write down some of the things you say and do. However, I will not write your name on the answer sheet so that no one else will know that they were your answers.

If you do not want to help, you do not have to. If you do not want to, I won’t mind and you won’t be in any trouble.

Also, if you agree to help but then change your mind, you can stop playing the game at any time you want, and don’t have to say why.

If you want to find out more about what I am doing, you can either talk to me or your teacher to find out more.

Do you want to help me today?

Yes  No

.....

Name of Participant

Date

.....

.....

.....

Name of Researcher

Date

Signature

---

**Office Use Only**

Participant’s Code: .....

Gender: M / F

D.o.B.: .....

## **APPENDIX C: Information Sheet.**

### **Notice of Proposed Research Project For Class X, Year Y**

#### **Information Sheet For Parents/Guardians.**



**Professional Doctorate in  
Clinical Psychology.**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Researcher:** Annie Pavitt

The children of [class X, year Y] are being invited to take part in a research project (named above). Before deciding whether you would like your child to take part, it may be useful for you to understand the rationale for this research and what it will involve.

#### **Background to the Study**

The aim of the research is to develop new ways of assessing children's abilities to problem solve through planning and carrying out actions. This is an important part of everyday life, yet there are few tests that assess this ability in children. It can be important to test this ability: for example if they are having problems at school, or have hurt their head in an accident. In this new, proposed procedure, we will be testing these abilities by using an adapted version of the traditional 'Guess Who?' game. In this study, we aim to discover whether a game-like procedure will engage children more than existing tests. This should help make these assessments more enjoyable and accurate.

#### **Do I have to allow my child take part?**

It is your choice whether or not to allow your child to be involved in the research study. If you decide that you do not want your child to participate, please fill in and return the slip at the end of this letter to the school within two weeks. You do not have to tell us why you do not want your child to take part. Additionally, you can withdraw your consent at any time (including after your child has taken part) by contacting the school or researcher (details overleaf). If you are happy for your child to take part, your child will be asked whether they want to take part. Before we start the task, it will be explained to your child that they do not have to take part, they will not be in any trouble if they do not want to, and that they can stop at any time without giving a reason.

**What will happen if I allow my child to take part?**

The researcher will discuss information about the study with your child. Your child will complete consent form if they want to join in. During the task, they be asked to play the adapted version of 'Guess Who?' and complete two other tasks. These tasks are measures which are already used in clinics, and are brief. One task involves spotting patterns, and the other task involves spotting similarities. All of the tasks combined should not take more than 25/30 minutes to complete. Your child will be given opportunities to ask questions before and after the tasks.

In the unlikely event that your child appears distressed, or to be significantly struggling during the task, they will be offered a break, and will be asked if they would like to reschedule or withdraw. In these circumstances, you will be notified about this.

**Are there any possible risks or disadvantages in taking part?**

There are no known risks or disadvantages in taking part in this study.

**What are the benefits of my child taking part?**

The results of this research project will provide information about how normally-developing children perform on this type of test. Ultimately, if this stage of the research is successful, we will be able to use this information to do further research with other children, in order to improve the way that we test certain children, for example, those who are seen in clinics who may have problems at school or illnesses that affect their executive functioning.

**Will my child's confidentiality be respected?**

All information and answers obtained within the study will be treated with the strictest confidentiality. The only people who will see your child's answers are the researcher and the project supervisor. Your child's consent form (which will have your child's name written on it), will be kept separate from the rest of the information.

**What will happen to the results from the study?**

The results of these tests will be assessed as a group, and therefore no information or feedback on individual children will be provided. The results will be written up for submission as a doctorate research project as part of a Clinical Psychology doctorate. Your child will not be identified in any part of the report.

**Who is this research organised by?**

The research is being conducted by Annie Pavitt as part of a doctorate in Clinical Psychology and is being supervised by Dr Matthew Jones-Chesters.

**Contact details of researchers:**

If you have any questions please email: X to contact Annie Pavitt, or X to contact Dr Matthew Jones-Chesters.

**APPENDIX D: Opt Out Form for Parents/Guardians.**

**Professional Doctorate in Clinical Psychology**

**Opt Out Form For Parents/Guardians.**



**Professional Doctorate in  
Clinical Psychology.**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Researcher:** Annie Pavitt

If you agree to your child taking part in the study if they want to, you do not need to take any further action.

However, if you do **NOT** wish your child to participate in the study as described above, please return the slip below.

Please return the slip before ..... or we will assume that permission is given.

**Researcher**

**Clinical Supervisor**

-----✂-----✂-----✂-----

**To:.....(name of form teacher)**

**I do NOT wish my child to participate in the research project.**

**Child's Name:** ..... **Class:** .....

**Parent/Guardian's Signature:** .....

**APPENDIX E: Consent Form – Parents/Guardians.**



**Professional Doctorate in Clinical Psychology  
Consent Form For Parents/Guardians.**

**Professional Doctorate in  
Clinical Psychology.**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Researcher: Annie Pavitt**

If you are happy for your child to take part, please sign below in order to say that:

- I have read the information sheet and understand the purpose of the study.
- I understand that my child's answers will be kept fully anonymous. I understand that any information which contains any person data for my child (including this sheet) will be kept in a secure location. My child's name, date of birth or gender will only be written on their consent form. For all other documents, my child will be given a code which will be allocated to them (which will be written on the bottom of this sheet by the researcher) if I wish to withdraw my consent at a later date, my child's answers can be matched with the consent form.
- I understand that my child will also be asked whether they want to take part, and therefore give consent, before taking part in the study.
- I understand that my child's participation is voluntary that that either they or I am free to withdraw their data from the project at any time, without giving a reason.
- I agree for my child to take part in this study if they want to.

Name of parent/guardian: .....

Name of child:.....

School:.....

Date:.....

**Office Use Only:**

Participant's Code: .....

**APPENDIX F: Head Teacher's *Loco Parentis* Form.**



**Professional Doctorate in Clinical Psychology  
Head Teacher's *Loco Parentis* Form**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Researcher:** Annie Pavitt

The study (title as above) has been fully explained to me. I have been given the opportunity to review the materials and ask questions.

The parents/guardians of the children who will be invited to participate in this study have been sent a letter home on *[date]* to inform them about the research.

Parents/guardians have been advised that they have a certain period of time (2 weeks) to withdraw (or 'opt-out') their child from participating in the study if they do not wish for them to take part.

I, as the head teacher of the school, am willing to act *in loco parentis* in giving my consent for the children (whose parents/guardians do not contact me) to participate in the study if they wish to.

.....

Name of head teacher

Name of school

.....

.....

Date

Signature

.....

.....

.....

Name of researcher

Date

Signature

**APPENDIX G:** Standardised Instructions for Child.



**Professional Doctorate in  
Clinical Psychology.**

**Project Title:** Using a game-like procedure as a new test of problem solving and concept formation in children.

**Standardised Instructions**

Today we are going to play a game. It is a new version of 'Guess Who?' with aliens. In front of you, you will see 24 pictures of aliens, which I would like you to take a moment to look at. Before the game started, I have already selected an alien which I would like you to guess. To do this, your task is to ask as few yes/no questions as possible in order to guess which alien I have chosen. Once you have guessed the correct alien, we will then play the game three more times.

Please do not discuss which aliens you had to guess with your friends, as this may change their answers.

Do you have any questions?

**APPENDIX H: Measure of Task Acceptability Using Likert/Visual Analogue Scale**



Professional Doctorate in  
Clinical Psychology.

*How much do you agree with this statement.....*

I enjoyed this task



1. Strongly Disagree



2. Disagree

3. Neither Agree or Disagree



4. Agree

5. Strongly Agree

-----

Office use only

Child's code: \_\_\_\_\_

Name of task: \_\_\_\_\_

**APPENDIX I: Coding Sheet**

	No. removed	No. left	Code
W6			
C: is it blue?			
M: no. you can put done ones that are blue.	9	15	1
C: is it yellow?			
M: no	7	8	1, 6
C: is it got yellow eyes?			
M: no	3	5	1
C: has it got a tail?			
M: yes	2	3	1
C: has it got 2 eyes?			
M: yes	1	2	1
C: has it got legs			
M: no			
C: is it fanf?	1	1	1
M: yes well done. I've got another one lined up///			
C: is it blue?			
M: no	9	15	1
C: has it got 2 eyes?			
M: no	8	7	1, 2
C: has it got 1 eyes?			
M: no	2	5	1, 6
C: is it grey?			
M: no	2	3	1

Figure 4. Example of a Coding Sheet Used by the Second Rater

Note: C= 'Child', M = 'Me' (researcher)

## APPENDIX J: Coding Scheme Explained

Table 16.

No.	Coding Category	Description	Examples
1.	'Constraint seeking'	Questions which attempt to eliminate two or more Aliens at the time asked (or a question attempting to eliminate one or more Aliens if four or less Aliens remain).	<i>Example:</i> Child: "Is it blue?" "Is it a triangle?" <i>Any question that eliminates more than two items, or one item if there are less than four items left overall.</i>
2.	'Hypothesis seeking'	These are questions which are <b>unrelated to the previous question</b> , when there was a <b>possibility of an 'Intradimensional' follow-up</b> to the previous question (see Code 6). For example, if a child asked "Is it blue?" and is told "no", the follows this up by asking "Does it have arms?", despite there being both yellow and grey Aliens left.	<i>Example:</i> Child: "Is it yellow?" Researcher: "No" Child: "Does it have wings?"  Categories that can follow this structure include: Colours, <i>Yellow, Blue, Grey</i> ; Shape, <i>Triangle, Square, Circle</i> ; Nose, Type <i>none, human, trunk</i> ; no. of antenna, <i>0, 1, 2</i> ; no. of eyes <i>1, 2, 3</i> ; no. of legs <i>0, 2, 4</i> ; skin, <i>scaled, fur, bald</i> .
3.	'Failure to eliminate'	Questions that did not eliminate any items.	<i>Example:</i> Child: "Is it blue?" Researcher: "Yes" Child: "Is it yellow?" <i>When there would be no yellow items standing.</i>
4.	'Delayed repetition'	A repeated a question from earlier in that trial.	<i>Example:</i> Child: "Is it a square?" Researcher: "No" Child: Is it a circle?"  Researcher: "No"

## APPENDIX J: Coding Scheme Explained - Continued

Researcher: "No"

			Child: "Is it a square?"
5.	'Immediate repetition'	Preservation; repeating the question just before it.	Child: "Is it a circle?" Researcher: "No" Child: "Is it a circle?"
6.	'Intradimensional'	Asking a question related to the question immediately before it. Therefore, following the same taxonomy (e.g. "Is the Alien a triangle?" "Is the Alien a circle?").	Child: "Is it yellow?" Researcher: "No" Child: "Is it blue?"  Categories that can follow this structure include: Colours, <i>Yellow, Blue, Grey</i> ; Shape, <i>Triangle, Square, Circle</i> ; Nose, Type <i>none, human, trunk</i> ; no. of antenna, <i>0, 1, 2</i> ; no. of eyes <i>1, 2, 3</i> ; no. of legs <i>0, 2, 4</i> ; skin, <i>scaled, fur, bald</i> .
7.	'Medium risk'	A question that would at a minimum eliminate a quarter or less of the options (i.e. asking a question that eliminates 6/24 items or less; see Appendix I for table demonstrating when to use this code). This is coded whether or not the question yielded a correct answer. This Code cannot be applied if there are only 10 or less Aliens left.	<i>Example:</i> Child: "Does it have four legs?" <i>This question would only eliminate 6 items at the start of the game.</i> <b>See Break Down for Code 7 (Appendix K)</b>
8.	'High risk'	A question which at a minimum eliminates a sixth or less of the items left (see Appendix I for table, regardless of whether the questions yields a correct response or not (e.g. a question related to 4/24 items or less). Do not rate when there are fewer than 5 aliens left.	<i>Example:</i> Child: "Does it have four legs?" <i>This question would only eliminate 4 items at the start of the game.</i> <b>See Break Down for Code 8 (Appendix K).</b>

**APPENDIX K: Scoring for Code 7 and Code 8**

Table 17.

Number of Aliens Left	Maximum Number of Aliens Removed by the Question to meet the codes	
	Code 7 'Medium risk'	Code 8 'High risk'
24	6	4
23	6	4
22	6	4
21	5	4
20	5	3
19	5	3
18	5	3
17	4	3
16	4	3
15	4	3
14	4	2
13	3	2
12	3	2
11	3	2
10	3	2
9		2
8		1
7		1
6		1
5		1

*Note: a question which eliminated of 5/24 on first question would be coded as Code 7 as it is less than 6 (but does not meet criteria for Code 8).*

**APPENDIX K: Scoring for Code 7 and Code 8 – Continued**

Table 18.

Number of Aliens Left	Maths to Work Out Minimum Numbers needed for Code 7 and Code 8 of Aliens Removed by the Question	
	Code 7 'Medium risk'	Code 8 'High risk'
24	$24 \div 4 = 6$	$24 \div 6 = 4$
23	$23 \div 4 = 5.75$ (rounded up to 6)	$23 \div 6 = 3.83$ (rounded up to 4)
22	$22 \div 4 = 5.5$ (rounded up to 6)	$22 \div 6 = 3.67$ rounded up to 4)
21	$21 \div 4 = 5.25$ (rounded down to 5)	$21 \div 6 = 3.5$ (rounded up to 4)
20	$20 \div 4 = 5$	$20 \div 6 = 3.33$ (rounded down to 3)
19	$19 \div 4 = 4.75$ (rounded up to 5)	$19 \div 6 = 3.17$ (rounded down to 3)
18	$18 \div 4 = 4.5$ (rounded up to 5)	$18 \div 6 = 3$
17	$17 \div 4 = 4.25$ (rounded down to 4)	$17 \div 6 = 2.83$ (rounded up to 3)
16	$16 \div 4 = 4$	$16 \div 6 = 2.67$ (rounded up to 3)
15	$15 \div 4 = 3.75$ (rounded up to 4)	$15 \div 6 = 2.5$ (rounded up to 3)
14	$14 \div 4 = 3.5$ (rounded up to 4)	$14 \div 6 = 2.25$ (rounded down to 2)
13	$13 \div 4 = 3.25$ (rounded down to 3)	$13 \div 6 = 2.17$ (rounded down to 2)
12	$12 \div 4 = 3$	$12 \div 6 = 2$
11	$11 \div 4 = 2.75$ (rounded up to 3)	$11 \div 6 = 1.83$ ( rounded up to 2)
10	$10 \div 4 = 2.5$ (rounded up to 3)	$10 \div 6 = 1.67$ (rounded up to 2)
9		$9 \div 6 = 1.5$ (rounded up to 2)
8		$8 \div 6 = 1.33$ (rounded down to 1)
7		$7 \div 6 = 1.17$ (rounded down to 1)
6		$6 \div 6 = 1$
5		$5 \div 6 = 0.83$ (rounded up to 1)

*Note: Code 7 is not to be coded when there are fewer than 10 Aliens left in the game. This was because the calculations for code 7 and 8 started to result in an overlap of results when there were fewer Aliens left. Code 8 should not be scored when there are fewer than 5 Aliens left.*

## APPENDIX L: Ethical Approval

### School of Psychology Research Ethics Committee

## NOTICE OF ETHICS REVIEW DECISION

For research involving human participants

**BSc/MSc/MA/Professional Doctorates in Clinical, Counselling  
and Educational Psychology**

**REVIEWER:** Rachel George

**SUPERVISOR:** Matthew Jones-Chesters

**COURSE:** Professional Doctorate in Clinical Psychology

**STUDENT:** Anne Pavitt

**TITLE OF PROPOSED STUDY:** Using a game-like procedure as a new test of problem solving and concept formation in children.

### DECISION OPTIONS:

1. **APPROVED:** Ethics approval for the above named research study has been granted from the date of approval (see end of this notice) to the date it is submitted for assessment/examination.
2. **APPROVED, BUT MINOR AMENDMENTS ARE REQUIRED BEFORE THE RESEARCH COMMENCES** (see Minor Amendments box below): In this circumstance, re-submission of an ethics application is not required but the student must confirm with their supervisor that all minor amendments have been made before the research commences. Students are to do this by filling in the confirmation box below when all amendments have been attended to and emailing a copy of this decision notice to her/his supervisor for their records. The supervisor will then forward the student's confirmation to the School for its records.
3. **NOT APPROVED, MAJOR AMENDMENTS AND RE-SUBMISSION REQUIRED** (see Major Amendments box below): In this circumstance, a revised ethics application must be submitted and approved before any research takes place. The revised application will be reviewed by the same reviewer. If in doubt, students should ask their supervisor for support in revising their ethics application.

**APPENDIX L – Ethical Approval - Continued**

**DECISION ON THE ABOVE-NAMED PROPOSED RESEARCH STUDY**

*(Please indicate the decision according to one of the 3 options above)*

**Minor amendments required** *(for reviewer):*

Check in relation to BPS current code and in relation to UEL SoP policy whether “opt out” consent is possible (as proposed here) and change to “opt in” if necessary.

**ASSESSMENT OF RISK TO RESEACHER** *(for reviewer)*

If the proposed research could expose the researcher to any of kind of emotional, physical or health and safety hazard? Please rate the degree of risk:

- HIGH**
- MEDIUM**
- LOW**

*Reviewer comments in relation to researcher risk (if any):*

**Reviewer** *(Typed name to act as signature):* R George

**Date:** 16.3.16

*This reviewer has assessed the ethics application for the named research study on behalf of the School of Psychology Research Ethics Committee*

**Confirmation of making the above minor amendments** *(for students):*

I have noted and made all the required minor amendments, as stated above, before starting my research and collecting data.

In the BPS code of human research ethics, section 10.1 states the following:

“In relation to the gaining of consent from children and young people in school or other institutional settings, where the research procedures are judged by a senior member of staff or other appropriate professional within the institution to fall within the range of

## **APPENDIX L: Ethical Approval - Continued**

usual curriculum or other institutional activities, and where a risk assessment has identified no significant risks, consent from the participants and the granting of approval and access from

a senior member of school staff legally responsible for such approval can be considered sufficient. Where these criteria are not met, it will be a matter of judgment as to the extent to which the difference between these criteria and the data gathering activities of the specific project warrants the seeking of parental consent from children under 16 years of age and young people of limited competence.”

Thus a head teacher can give permission and act in *Loco Parentis* for participants if the activities are not unusual and there is low risk. For each school, the school's own policy regarding consent will be followed.

Student's name (*Typed name to act as signature*): ANNIE PAVITT

Student number: U0500741

Date: 29<sup>th</sup> April 2016

*(Please submit a copy of this decision letter to your supervisor with this box completed, if minor amendments to your ethics application are required)*

### **PLEASE NOTE:**

\*For the researcher and participants involved in the above named study to be covered by UEL's insurance and indemnity policy, prior ethics approval from the School of Psychology (acting on behalf of the UEL Research Ethics Committee), and confirmation from students where minor amendments were required, must be obtained before any research takes place.

\*For the researcher and participants involved in the above named study to be covered by UEL's insurance and indemnity policy, travel approval from UEL (not the School of Psychology) must be gained if a researcher intends to travel overseas to collect data, even if this involves the researcher travelling to his/her home country to conduct the research. Application details can be found here:

<http://www.uel.ac.uk/gradschool/ethics/fieldwork/>

**APPENDIX M: Mean Learning Slope Total and in Different Age Groups**

Table 19.

	Participant	Trial 4 AS	-	Trial 1 AS	= Learning Slope
7-8 year olds	1	16		13	3
	2	17		18	- 1
	3	14		12	2
	4	16		17	- 1
9-10 year olds	5	19		18	1
	6	15		15	0
	7	11		11	0
	8	7		16	- 9
	9	17		19	- 2
	10	17		19	- 2
	11	14		10	4
11 year olds	12	17		18	-1
	13	12		15	-3
	14	15		13	-2
	15	18		16	2
	16	13		17	-4
	17	16		13	3
	18	12		19	-7