USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN

Emily Hay

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

2023

ABSTRACT

Concept Formation (CF) is an important skill necessary for academic performance, everyday functioning, and lifelong achievements (Blair & Razza, 2007). CF difficulties have been associated with social disadvantage possibly due to reduced learning opportunities (Blair, 2002). CF difficulties have also been associated with certain neurodevelopmental disorders, such as autism-spectrum disorder (Kim et al., 2020). Importantly, research has shown that early intervention can support development of CF skills and improve quality of life (Pasnak, 2006). Despite its importance there are few assessments suitable for children, and assessments available in the UK often advantage English-speaking examinees who are familiar with Western culture.

The Alien Game was piloted by Pavitt (2017) and evidence was found to support its utility as a more culturally fair and child-friendly assessment of CF. This study aimed to further develop the Alien Game by improving the stimuli, developing a scoring procedure, and exploring concurrent and predictive validity.

Thirty-four children aged 8-11 years took part in this second pilot study, and five measures of performance in the Alien Game were identified. Performance was not found to correlate with verbal abilities, and speaking English as a first language did not support performance. A significant correlation was found between WISC-IV Matrix Reasoning (Wechsler, 2003) and performance, providing evidence of concurrent validity. No relationship was found between performance in the Alien Game and scores obtained via teacher-rated CHEXI (Thorell & Nyberg, 2008), so predictive validity could not be established. Importantly, children appeared to be engaged with the game and provided positive feedback.

Overall, these findings provide support for the Alien Game as a more culturally-fair and child-friendly measure of CF, which has clinical implications as an affordable and easy to administer screening tool. Early identification of CF difficulties can allow for targeted learning plans to be established to support children to develop this important skill.

ACKNOWLEDGEMENTS

Firstly, I would like to thank all the children who participated in this study for their time and invaluable feedback which allowed us to complete this research project. I would also like to thank the headteacher of the school for providing us with the opportunity to run the study in their school, and to the class teacher for being so patient and for completing the questionnaires. I would also like to say a big thank you to the school SENDCo for supporting us to organise data collection, and for being so helpful throughout.

Thank you to my supervisor, Dr Matthew Jones-Chesters, for his advice and guidance throughout this project, and for proving feedback on the draft of my thesis. I would also like to thank my thesis group for their support and sharing of ideas during the development of this project.

Finally, I would like to thank my friends for all their support, and for providing fun and laughter during stressful times, and a special thanks to Tony for his constant support and encouragement throughout the last three years.

TABLE OF CONTENTS

1. INTRODUCTION ... 8
1.1. Executive Functions ... 8
1.1.1. Concept Formation ... 9
1.1.2. Development of Concept Formation Abilities ... 10
1.1.3. Executive Functions and Socioeconomic Status ... 11
1.1.4. Executive Functions and Academic Attainment ... 12
1.1.5. Executive Functions and Neurodevelopmental Disorders ... 13
1.1.6. Interventions for Concept Formation ... 14
1.1.7. Summary of Section ... 15
1.2. Assessment of Concept formation Abilities ... 16
1.2.1. Format ... 16
1.2.2. Available Tests of Concept Formation ... 17
1.2.3. Problems with Current Tests of Concept Formation ... 18
1.2.4. Summary of Section ... 22
1.3. Games As a Source of Engagement ... 22
1.4. Literature Review ... 24
1.4.1. Methods of Literature Review ... 24
1.4.2. Search Strategy ... 24
1.4.3. Inclusion and Exclusion Criteria ... 24
1.4.4. Search results ... 25
1.5. Review of Literature ... 26
1.5.1. The Tinker Toy Test ... 26
1.5.2. The Object Classification Task for Children (OCTC) ... 27
1.5.3. The Flexible Item Selection Task (FIST) ... 28
1.5.4. Prototypicality Ratings and Conceptual Development ... 30
1.5.5. A novel measure of matching categories for early development ... 31
1.5.6. The Robots Task ... 32
1.5.7. Summary of Literature Review ... 34
1.6. The Alien Game ... 35
1.7. Current Study ... 36
1.7.1. Research Aims ... 36
1.7.2. Research Questions ... 37
2. EPISTEMOLOGY & METHODOLOGY ... 38

2.1. Epistemology 38
2.2. Design 39
2.3. Materials 39
2.4. Development of The Alien Game 40
2.4.1. Illustrations 41
2.4.2. Game Apparatus 43
2.4.3. Game Procedure 43
2.4.4. Summary of Development of the Alien Game 43
2.5. Inclusion & Exclusion Criteria 44
2.6. Ethics 45
2.7. Confidentiality 46
2.8. Procedure 46
2.9. Data Collection and Analysis 47
2.10. Scoring Procedure 48
2.10.1. Total Questions 49
2.10.2. Weighted Achievement Score (WA) 49
2.10.3. Abstraction Score (Ab) 50
2.10.4. Initial Abstraction Score (IA) 52
2.10.5. Ineffective or Unallowed Questions (IU) 53
2.10.6. Alien Game Score (AG) 55
2.10.7. Summary of Scoring Procedure 56
2.11. Alien Game Evaluation 56
2.11.1. Acceptability 56
2.11.2. Evaluation of Attributes 57
2.12. Participants 57
2.12.1. Cognitive Data 60
3. RESULTS 63
3.1. Exploratory Data Analysis 63
3.2. Alien Game Evaluation 63
3.2.1. Acceptability 63
3.2.2. Evaluation of Attributes 64
3.3. Alien Game Performance 65
3.3.1. Total Questions (TQ) 66
3.3.2. Weighted Achievement Score (WA) 66

3.3.3. Initial Abstraction Score (IA) ... 67
3.3.4. Abstraction Score (Ab) ... 67
3.3.5. Ineffective or Unallowed Questions (IU) Score ... 68
3.3.6. Weighted Ineffective or Unallowed Questions (WIU) Score ... 68
3.3.7. Alien Game (AG) Score ... 68
3.3.8. Summary ... 69
3.4. Validity ... 69
3.4.1. Correlation Between Alien Game Measures ... 69
3.4.2. Correlation Between Concept Formation Measures ... 70
3.4.3. Concurrent validity ... 71
3.4.4. Predictive validity ... 71
3.4.5. Influence of demographic data on Alien Game performance ... 71
3.4.6. Validity summary ... 73
3.5. Reliability ... 73
4. DISCUSSION ... 75
4.1. Summary of Research Questions and Aims ... 75
4.2. Summary of Research Findings ... 75
4.2.1. Research Question One ... 75
4.2.2. Research Question Two ... 79
4.2.3. Research Question Three ... 80
4.2.4. Research Question Four ... 80
4.2.5. Summary of Findings ... 81
4.3. Exploration of the Scoring Procedure ... 81
4.3.1. Total Questions ... 81
4.3.2. Weighted Achievement Score ... 82
4.3.3. Ineffective or Unallowed questions ... 82
4.3.4. Abstraction Score ... 83
4.3.5. Alien Game Score and Weighted Ineffective or Unallowed questions ... 84
4.4. Concept Formation ... 84
4.4.1. Relationship With Other Executive Functions ... 84
4.4.2. Learning Conceptual Skills ... 86
4.4.3. Assessment of Concept Formation ... 86
4.5. Critical Evaluation ... 89
4.5.1. Strengths ... 89

4.5.2. Limitations ... 90
4.6. Clinical Implications ... 92
4.7. Future Directions ... 93
4.7.1. Further Development of the Alien Game Stimuli ... 93
4.7.2. Further Development of the Scoring Procedure ... 93
4.7.3. Reliability ... 94
4.7.4. Generalisability ... 94
4.8. Concluding Summary ... 95
References ... 96
APPENDICES: ... 113
APPENDIX A: Pictures used in Pavitt's (2017) Alien Game ... 114
APPENDIX B: Pictures Used in the Alien Game. ... 115
APPENDIX C: Participant Information Sheet -Child ... 116
APPENDIX D: Participant Consent Form - Child ... 118
APPENDIX E: Participant Debrief Form - Child ... 119
APPENdIX F: Participant Demographic Record Form ... 120
APPENDIX G: Alien Game Response Record Form ... 121
APPENDIX H: Likert/visual Analogue Scale as Measure of Task Acceptability ... 122
APPENDIX I: CHEXI questionnaire (Thorell & Nyberg, 2008) ... 123
APPENDIX J: Study Advertisement ... 124
APPENDIX K: Parental / Guardian Information Sheet ... 125
APPENDIX L: Parental / Guardian Opt Out Consent Form ... 129
APPENDIX M: Parental Debrief Form ... 131
APPENDIX N: Loco Parentis Consent Form ... 133
APPENDIX O: Standardised Instructions for the Alien Game ... 134
APPENDIX P: Copy of Ethics Application ... 135
APPENDIX Q: UEL Risk Assessment Form ... 145
APPENDIX R: Notice of Ethics Review Decision Letter ... 148
APPENDIX S: SPSS Output Histograms and Boxplots ... 152

LIST OF TABLES AND FIGURES

Table 1. Weighted Achievement Score Calculation……………………………………50
Table 2. Total Number of Ineffective or Unallowed Questions………………………. 54
Table 3. Weighted Ineffective Questions (WIU's) scoring…………………………....55
Table 4. Participant Ethnicities…………….…………………………………….…….58
Table 5. Participant Primary Language Spoken at Home……………………………..59
Table 6. Participant Fathers Occupation………………..……………………………60
Table 7. WISC-IV and CHEXI age-scaled scores, and group differences based on Gender and English as a Primary Language……………………..……61
Table 8. Likert scale enjoyability rating for the Alien Game…………………………..64
Table 9. Frequency and percentage of questions asked about each alien attribute………...………………………………………………………………65
Table 10. Descriptive Statistics for Novel Alien Game Measures……………………66
Table 11. Spearman's Rank Correlations within the Alien Game Measures………..70
Table 12. Spearman's Rank Correlations between Concept Formation Measures…………………………………………………….…………70
Table 13. Spearman's Rank Correlations between Alien Game Measures, WISC-IV and CHEXI subscales…………………………………….……….71
Table 14. Alien Game Performances Grouped by Age, Sex and English as Primary Language………………………………………………………...72
Table 15. Descriptive Statistics and Wilcoxon Signed Ranks Test comparing Performance Measures across Trials…………………….…………………...74

Figure 1. PRISMA (2009) Flow Diagram of Article Selection Process………………25
Figure 2. Example of Illustration Development: Ears and Wings……………………41
Figure 3. Example of Illustration Development: Skin texture and Eye Colour……..42
Figure 4. Ethnic Composition in Sample Compared to London……………………..58

1. INTRODUCTION

The purpose of this research is to investigate the assessment of concept formation (CF) in children. CF is an executive function which has been directly associated with academic performance, everyday functioning, and lifelong achievements (Blair & Razza, 2007), though despite the importance, there has been a lack of suitable paediatric assessments of CF. Due to this, Pavitt (2017) developed a test of CF for children using a game-like approach to produce an assessment which is more child-friendly, and conducted a pilot study to assess its potential. This study aims to further develop this assessment, and to conduct a second pilot study.

This section will introduce the area of research, beginning by discussing the importance of executive functions and CF skills, and then will briefly discuss the current tests of CF, and the issues of using these with a paediatric population. Finally, a literature review will be presented which will position the current study within the existing literature, which will lead to the research aims and questions.

1.1. Executive Functions

The term “executive function” (EF) is an umbrella term used to describe a set of top-down cognitive processes such as working memory, mental flexibility, and inhibition (Diamond, 2013; Lucenet & Blaye, 2014). These skills are thought to be “core” EFs and are the building blocks for “higher order” EFs, such as planning, reasoning, and problem solving. EFs are mediated by the frontal lobes, specifically the dorsolateral prefrontal cortex and connections to wider neural networks (Curtis & D’Esposito, 2003). The developmental trajectory of EF skills is unique to each individual (Welsh et al., 1991) and have been found to correlate with maturation of neural connections in the prefrontal cortex, basal ganglia, and parietal lobes through increased neuronal myelination, synaptic pruning, and consolidation of synapses through experience (Klingberg et al., 1999).

EF skills are essential for social and psychological development as well as physical and mental health (Diamond, 2013), and have been found to predict school attainment better than measures of general intellectual ability, such as ‘IQ’ (Blair & Razza, 2007), as well as adult socioeconomic status and likelihood of criminal conviction (Moffitt et al., 2011). These skills are necessary for day-to-day tasks, including decision making, setting goals, and actioning plans (Lezak, 1982).

1.1.1. Concept Formation

CF is a higher-order EF and may be defined as the process by which an individual learns to categorise experiences and objects according to rules (Alt et al., 2013). To form a concept, one must distinguish relevant from irrelevant stimuli, abstract a rule, pattern, or relationship, and then be able to generalise this to novel situations (Smidts et al., 2004). CF therefore involves abstraction skills (the ability to place examples into categories), and induction skills (the ability to identify patterns and rules).

Concepts can be concrete and perceptual (e.g., colour, shape, or size) or abstract and conceptual (e.g., ‘a leader’); categorizing items or experiences into these more abstract classes is harder as members may not relate completely (Smidts et al., 2004). Children, therefore, often use more concrete methods to categorise compared to adults, as abstract thought develops with age and with exposure to learning opportunities (Piaget, 1952). In categorisation tasks, children may over or under generalise categorisation rules, allow for less variability in category membership, and be less forgiving of atypical examples (Alt et al., 2013).

Conceptual knowledge can be argued as the foundation of intellectual ability (Kagan, 1966; Mpofu et al., 2017), as the ability to categorise is a building block to more advanced skills (Condy et al., 2021). Advanced conceptual skills include the ability to understand conservation of liquids, cause and effect, force and motion, as well as mathematical notions such as magnitude of numbers and operations (Göksun et al., 2013). Conceptual knowledge is also necessary for language development such as understanding linguistic concepts, grammatical rules, and word learning (Condy et al., 2021).

CF is also a skill required to perform well in traditional tests of IQ, and conceptual knowledge is essential for understanding administrative instructions in neuropsychological testing (Condy et al., 2021). The importance of being able to accurately assess these skills is therefore relevant to all children (Mpofu et al., 2017).

1.1.2. Development of Concept Formation Abilities

Piaget (1952) held a constructivist view of conceptual development and believed such skills and abilities develop throughout childhood, supported and enhanced by one's environment. Piaget's theory of cognitive development (1952) suggests that a child's conceptual skills are constrained by their age. He suggests that children under 7 (*preoperational* stage) do not have capacity for conceptual thinking, but these skills develop rapidly between ages 7-8 years and 11-12 years. Before age 11 years (*concrete operational* stage), conceptual thinking abilities are limited to concrete categories rather than abstract thought (Freyberg, 1966), and according to Piaget, it is not until after 12 years of age (*formal operational* stage) that abstract thought develops. Piaget's (1952) theory also provides an account of how concepts are organised. He suggested that children develop a framework to understand and organise the world and knowledge (a 'schema'). As the child encounters new knowledge through their interaction with their environment, adaptation occurs where these schemas update.

While Piaget's theory offers a framework to understand the development CF, a few assumptions have been disputed. Firstly, it has been suggested that Piaget's tasks were too difficult for children and were not sensitive enough to identify conceptual thought in young children. Additionally, his approach to testing has been described as 'adultcentric' by failing to engage young participants (Matusov & Hayes, 2000), and more recent attempts to replicate his experiments with child-friendly adaptations have yielded different results (Borke, 1975).

Additionally, Piaget's theory has been described as universalist and ethnocentric due to the generalisation of Western middle-class experiences to non-Western cultures, and the failure to attend to cultural differences and social influences on cognitive development (Matusov & Hayes, 2000; Sanghvi, 2020). Piaget, however, did argue that abstract thought in the *formal operation* stage may depend on specific experiences of Western schooling such as hypothesis-testing in science lessons; so constructs such as conservation of liquid may not be understood by a child who has not been specifically taught this notion (Piaget, 1995).

Mpofu et al. (2017) suggests that conceptual development is important universally, with many concepts relevant across languages and cultures (e.g., social and emotional concepts, perceptual concepts including shape, size and colour, and concepts of time, sequence, and directions). While CF may be universally important, studies have demonstrated differences in how information is understood and organised across different cultures (Micheals & Cazden, 1986). Variation in CF development can also be found within the same culture due to the importance of a stimulating environment with rich learning opportunities in the development of EF skills. A lack of such opportunities, such as experienced by children from disadvantaged families, can impair EF development (Ford et al., 2019).

1.1.3. Executive Functions and Socioeconomic Status

Socioeconomic status (SES), used to measure social advantage, has been found to predict childhood EF skills better than other cognitive abilities (Blair, 2002; Farah et al., 2006; Lawson et al., 2018). SES is often measured via parental education, occupation, and income (Lawson et al., 2018) and refers to the availability of resources, both economic (e.g., material wealth and income) and social (e.g., education and status; Fitzpatrick et al., 2014).

Low SES, poverty, malnourishment, illness, and lack of access to healthcare have been found to negatively impact cognitive functioning (McCoy et al., 2015), verbal development (Fernald et al., 2013) and EFs (Ford et al., 2019). SES is associated with quantity and frequency of stressful life events, as well as quality of parenting and schooling (Duncan & Magnuson, 2012), which may be mediating factors in the relationship between SES and cognitive development.

The relationship between SES and EFs may be as a result of reduced frequency and richness of learning opportunities and reduced quality of caregiver input available to disadvantaged children as a result of parental stress, which are factors known to foster strong EF skills (Fitzpatrick et al., 2014; Ford et al., 2019; Johnson et al., 2016). Associations have been found between extreme adversity, such as abuse and psychosocial deprivation, with less developed neural systems particularly within the frontal lobes (Lawson et al., 2018; Sheridan et al., 2012). These findings, however, may be unique to the US and UK where much of this research has taken place, and may not be the same across the globe where environmental factors vary significantly. Within the UK, however, SES has also been found to impact other areas of a child's development, such as their social relationships and academic attainment (Fitzpatrick et al., 2014).

1.1.4. Executive Functions and Academic Attainment

Some studies which have identified a relationship between SES and academic attainment, suggest that this relationship may be mediated by verbal skills (Noble et al., 2005), while others have found EF to mediate this relationship even when verbal abilities (Dilworth-Bart, 2012) and IQ (Fitzpatrick et al., 2014) are controlled for. Strength of EF skills in childhood is linked to learning, academic performance, and adult career attainment (Blair & Razza, 2007; Heckman et al., 2006), and conceptual knowledge has been strongly associated with scores on tests of general intelligence (McIntosh et al., 1995; Mpofu et al., 2017). Maths, English, and science ability in particular have been associated with conceptual skills, as the ability to form concepts is essential for understanding classroom teaching material (Bracken, 1986), and improvement in EFs has been found to improve school grades (Finn et al., 2014; Holmes et al., 2010).

In a study investigating the relationship between development of conceptual skills and academic attainment, Freyberg (1966) found a stronger correlation between conceptual ability and mental age than chronological age. It was also found that conceptual skills accounted for a statistically significant proportion of variance in arithmetic abilities, and scores on tests of conceptual knowledge were found to be stronger predictors of academic attainment than intelligence test scores.

In addition to conceptual knowledge, behaviour regulation may also play a role in the relationship between EF and academic attainment. Classroom behaviour influences the learning environment, such as the ability to follow instruction, pay attention, and ignore distractions (Lin et al., 2003). EF difficulties can also result in poor emotion regulation which can result in challenging behaviours in the classroom and in adulthood (Eslinger et al., 1992; Moffitt et al., 2011).

While many studies have identified a direct link between EF and academic attainment, Fitzpatrick et al. (2014) also found SES to predict academic achievement when EF and IQ were controlled for. Previous research has suggested that children from disadvantaged families may be less engaged in learning and be less task-oriented, possibly as a direct consequence of stress (Duncan et al., 2007). It is important to understand the factors contributing to disparities in academic attainment to develop targeted learning plans. For example, conceptual skills can be taught by parents in the home, and by teachers across all year groups, and doing so has been shown to significantly improve educational gains (Wilson, 2004). Furthermore, a number of interventions have been developed to improve EF with the aim to reduce SES inequalities in academic attainment (Diamond & Lee, 2011; Neville et al., 2013). These studies suggest that interventions directed at developing EF's may promote reading and maths skills, especially for disadvantaged children.

1.1.5. Executive Functions and Neurodevelopmental Disorders

Neurodevelopmental disorders such as attention-deficit-hyperactivity disorder (ADHD), autism-spectrum disorder (ASD) and learning disabilities (LD) (Watson, 2016) have also been associated with difficulties with EFs (Kim et al., 2020). Children with these diagnoses have also been found to have less developed conceptual knowledge, possibly related to language delays. Language skills are related to the development of conceptual skills (Yoshida & Smith, 2005), and many cognitive assessments rely on verbal responses. As a result cognitive skills such as conceptual knowledge can be difficult to assess in those with limited verbal abilities despite the importance of doing so (Alt et al., 2013).

Children with neurodevelopmental disorders may also struggle with conceptual skills, which is important for effective information processing (Rosch, 1978). Alderson-Day and McGonigle-Chalmers (2011) suggest that children with ASD may experience difficulties with categorisation, especially when this shifts from concrete information to more abstract or complex materials (Ropar & Peebles, 2007). Dysexecutive difficulties in children with ADHD have been associated with difficulties learning, academic performance, and behavioural difficulties (Schreiber et al., 2014). Diagnostic assessments of neurodevelopmental disorders are time consuming and resource heavy (Putra et al., 2020). Early screening of some of these difficulties independent of a diagnostic assessment can therefore allow for early intervention to be provided to target specific difficulties and improve quality of life (Kim et al., 2020).

1.1.6. Interventions for Concept Formation

Researchers have called for educators to support children to develop conceptual skills from an early age in order to support overall cognitive development, as well as general and emotional intelligence (Mpofu et al., 2017). Tzuriel and Klein (1985) studied children’s ability to learn concepts and developed a measure of children’s cognitive modifiability called The Children's Analogical Thinking Modifiability (CATM) test. This test was used to measure CF abilities before and after an intensive training period where children were taught how to complete analogical puzzles. Results showed a significant improvement in score from pre-test to post-test for all groups following the intervention. This improvement was largest for disadvantaged children, who moved from intermediate performance in the pre-test, to performing better than all other groups following the intervention. Children assigned to a special education group and additional needs group also made gains, though these gains were smaller. These findings suggest that all children can learn these conceptual skills, and that disadvantaged children may benefit the most from this.

Pasnak and colleagues have investigated whether conceptual skills can be taught and improved using the “oddity principle” which is the ability to identify which object within an array is the odd one out (Pasnak et al., 1991, 1996). Pasnak et al. (2006) studied whether improvements in the oddity principle can lead to improvements in literacy for children attending Head Start programmes in Northern Virginia. The Head Start programme is for children from low-income families who are deemed at risk of experiencing difficulties at school. Their task involved viewing a series of four common items, three of which were similar in one dimension, while the other differed by shape, size, orientation, or conceptual category. A game-like element was added to make the task more engaging, by giving the children a pony or a dinosaur to use to identify their answer. These games were played daily with the children from October until May.

Results found that 17 of the 20 children mastered all games by the end of the learning period and those children also improved their performance in other tests of the oddity principle. This suggests that once learnt, children could generalise this skill to new concepts, formats, and contexts. Those in the experimental condition also scored higher on numeracy tests, suggesting that interventions aimed at improving conceptual skills may support children’s academic attainment.

Additional evidence for this was found by Bottino et al. (2007) who attempted to utilise popular brainteaser games to measure and improve logical and strategic reasoning, critical and reflective thinking, and EF skills. Following the intervention, the classes included in this study performed better than equivalent age group classes at the same school in standardised tests of language, science, and maths, suggesting that playing these games can improve children’s academic performance. Furthermore, school programs targeting EFs, such as the Perry Preschool Programme (Almlund et al., 2011) have been found to promote academic attainment. This programme focused on fostering skills in planning, active learning, self-regulation, and self-control, and while sustained improvement in IQ were not found, lasting improvements in school attainment and adulthood adjustment were identified.

1.1.7. Summary of Section

In summary, CF is an important skill for all children and the development of this skill has been associated with environmental factors such as exposure to rich learning opportunities. Children from disadvantaged backgrounds have been found to have less developed EF skills, possibly due to parental stress and reduced parental input. The relationship between SES and academic attainment has been found to be mediated by EF skills which are crucial for learning. It is therefore important to have available assessments of these skills, so specific learning plans can be developed, as there is evidence that interventions targeted at developing conceptual skills can support learning.

1.2. Assessment of Concept formation Abilities

The following sections will now discuss the tests currently available to assess CF. There is a large contemporary literature base investigating expressive executive functions which are required for goal-directed behaviour, such as task setting, task switching, and inhibition. There are also established paediatric assessments of these expressive skills including the Test of Everyday Attention for Children (TEA-Ch; Manly et al., 1999) and The Developmental Neuropsychological Assessment (NEPSY; Korkman et al., 1998). There is far less research and fewer available assessments, however, for receptive executive functions, such as CF and abstraction. While a comprehensive review of the current available tests of CF is beyond the scope of this thesis, the following section will provide a brief overview.

1.2.1. Format

Tests of CF are either organised as single-trial format, where all information needed to complete the task is provided within the trial (for example, in WISC-IV Similarities or Matrix Reasoning (Wechsler, 2003), each trial is independent of the previous question and answer) or multi-trial format where information to complete the overall task is gathered over a series of trials (for example, in the DKEFS 20-Questions task; Delis et al., 2001), the question asked by the examinee depends on the answer to the previous question; Pavitt, 2017).

Multi-trial tasks require multiple cognitive demands in addition to CF skills, such as working memory (Smidts et al., 2004). While this may result in difficulties identifying a particular area of concern, it offers more ecological validity as such skills do not exist in isolation in day-to-day life. Single-trial tasks on the other hand allows for a specific skill to be tested in isolation (Pavitt, 2017).

CF can also be assessed visually or verbally. Verbal tests of CF require semantic and linguistic knowledge and so such tests are not suitable for individuals with receptive or expressive language difficulties or for very young children who are at early stages of language development (Alt et al., 2013). Conceptual knowledge underlies verbal skills so those who have conceptual difficulties are likely to also have communication difficulties, making such difficulties challenging to distinguish

(Tecoulesco et al., 2021). CF is an important skill to measure in children with language difficulties to develop targeted treatment plans, so visual tests may be more suitable (Alt et al., 2013).

Additionally, tests developed in Western countries are often available only in English language making them inaccessible to many seeking neuropsychological assessments in the UK for whom English is not their first language. Furthermore, such tests often use Western concepts and therefore lack construct validity when translated (see section 1.2.3.1 for more detail). To produce a more culturally fair test, therefore, the need for verbal skills should be minimised.

1.2.2. Available Tests of Concept Formation

The Similarities task in the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V, Wechsler, 2014) is an example a verbal single-trial test of CF in which the examinee is presented with two concepts (for example, “butterfly and bee”, or “revenge and forgiveness”) and asked to identify a relationship between them. As a single-trial test, this task minimises demands on working memory, though relies on understanding of English language and familiarity with Western concepts (see section 1.2.3.1).

An example of a verbal multi-trial test of CF is the D-KEFS 20-Questions task (Delis et al., 2001) which is suitable for children aged 8 years and over. In this task, participants are shown an array of images of items such as vehicles, animals, or plants, and participants are required to identify the target item by asking a series of yes or no questions. This test requires working memory and language abilities in addition to conceptual skills, though it has been claimed to be more enjoyable task due to its game-like format (Gioia, 2015).

The Cattel Culture-Fair Intelligence Test (Cattell & Cattell, 1973) is an example of a visual single trial test of non-verbal intelligence, which assesses CF skills. The task includes a set of matrices and participants are required to find the missing element of a visual puzzle. This test is available for children ages 4+ and was developed as a culturally fair and unbiased test due to the visual nature of the task, however this assumption has been questioned (Nenty & Dinero, 1981).

The Wisconsin Card Sorting Test (WCST, Heaton et al., 1993) is a visual multi-trial test of CF for children aged 7 years and over. In this test, examinees are instructed to sort cards into categories based on perceptual dimensions such as shape, colour, or number, and the rule regarding which perceptual dimension to sort by changes without warning. Participants are therefore required to abstract the rule, conceptualise categories, inhibit irrelevant information, shift rules, and use working memory, and so it can be difficult to isolate the problem area for those who have difficulties completing this task (Jacques & Zelazo, 2001).

1.2.3. Problems with Current Tests of Concept Formation

1.2.3.1. *Culture*: Neuropsychological assessments available in the UK often advantage English speaking examinees who are familiar with Western culture. Culture can be defined as the way a group of people live, including shared ways of behaving, feeling, and thinking, as well as shared knowledge, beliefs, and attitudes (Rosselli & Ardila, 2003). Cultural differences include aspects of identify such as nationality, ethnicity, language, and religion (Van de Vijver, 1997).

The culture in which one lives determines and shapes the way a person thinks and understands the world (Vygotsky, 1978). Differences in cultural and social background is an important contributor to variance in neuropsychological assessment (Nell, 2000), which was demonstrated by Luria (1979) who found that different social experiences, including education and social background, resulted in different approaches in cognitive tasks.

Cognitive assessments which have been developed in the West are often not suitable for all individuals who may require them. Most available neuropsychological assessments in the UK are only available in English, and often involve complex instructions and require verbal responses from examinees. Tests are sometimes translated from English to other languages with the aim to make them more accessible to non-English speakers; however, such tests often involve constructs that are Western in origin and may not be relevant to the culture the examinee is familiar with. This may result in a loss of construct validity, and may therefore lead to an inaccurate representation of the individual’s abilities (Haddlesey, 2016).

Additionally, many CF assessments are based on models of intelligence such as Spearman's 'g factor' (Spearman, 1904), Wechsler's full-scale intelligence quotient (Wechsler, 1949) and Cattell's theory of fluid vs. crystallized intelligence (Cattell, 1963). These understandings of intelligence, however, are not universal, and may differ between cultures (Mpofu et al., 2017). Furthermore, Western cultures value individual achievement, and education is organised around exams so children who attend formal education in the West learn how to perform well in tests and thus have an advantage in neuropsychological assessments (Haddlesey, 2016). Literacy skills as measured in these tests are also confounded by the assumption of shared knowledge such as particular animals, buildings, and weather, which differ across the world (Ford et al., 2019).

All these issues discussed so far can lead to diagnostic errors and false positives in neuropsychological assessment outcomes (Norman et al., 2011) which underlies the need for culturally fair tests with normative data reflective of a multicultural society. To achieve culturally fair tests, verbal requirements should be minimised, both in task instruction and response, and to refrain from using concepts and notions that are culturally specific. Additionally, tests that have available norms based on a representative sample of the population they are serving, are also more culturally fair. In the past, attempts have been made to produce culturally fair cognitive assessments through use of nonverbal measures to reduce language requirements. However, van de Vijver (1997) argues that less attention has been made to the materials and procedure used. For example, tasks such as drawing maps or copying figures, which are skills practiced in Western schooling, may disadvantage individuals from some cultures where such skills are not regularly used (Ardila & Moreno, 2001). Additionally, the use of nonverbal stimuli is not sufficient alone for a test to be considered culturally fair (Rosselli & Ardila, 2003). Fernandez and Abe (2018) argue that in order to create a test which is valid cross-culturally, both the content of the assessment, and the cognitive function being tested, should be meaningful cross-culturally.

Van de Vijver (1997) identified ways in which bias occurs when Western neuropsychological assessments are administered to individuals from non-Western cultures. Firstly, the authors argue that bias occurs when differences in test performance occurs due to factors other than differences in the skill being measured. They suggest this may be a result of construct bias, method bias, or item bias. Construct bias occurs when the underlying cognitive function being measured is not equivalent cross-culturally, for example, intelligence, which has been defined by Western standards and is not a universal construct. Method bias originates from the instrument itself, for example when familiarity with the test materials and stimuli leads to group differences in performance, or when communication difficulties between the examiner and examinee lead to poor performance due to misunderstanding of what is required. Lastly, item bias refers to specific items included in the test differing on applicability between cultures, for example, the use of concepts or pictures which are culturally laden.

Van de Vijver and Tanzer (2004) propose several methods to address these problems. Firstly, they suggest that construct bias can be avoided by 'decentering' which involves removing words and concepts which are culturally specific, piloting the assessment within the culture it is being used, and then making necessary modifications. Additionally, method bias can be avoided by developing detailed instructions and testing protocols so those familiar with the instrument are not advantaged, and by including practice trials to improve familiarisation with the testing materials and response protocol. Item bias can be avoided by assessing each item to remove culturally laden constructs or using psychometric assessment methods to detect item bias. Importantly, the examiner having a thorough understanding of how culture can impact test performance allows them to take a critical approach when scoring and interpreting neuropsychological assessment data cross-culturally.

Fernandez and Abe (2018) additionally propose that the development of new tests based on current knowledge of how culture impacts neuropsychological testing can address these issues. They argue that such test should involve culturally fair items, content, and response format, and should use stimuli that is less likely to be influenced by Western education, such as the use of everyday objects rather than literacy and mathematical operations.

1.2.3.2. *Engagement:* Most commonly used neuropsychological assessments of CF were originally designed for adults, and while many now have established formats for children, they may not be suitable (Kim et al., 2020). EFs develop at unique trajectories throughout childhood, and abilities change rapidly over a relatively short space of time (Blair, 2016). Tests designed for adults do not take the developmental level of the child into consideration, such as reading and literacy skills (Berg et al., 2020). Assessments aimed at children therefore require adaptations across the age range, in order to meet the developmental needs of children and to avoid floor effects (when a test is too challenging for the child's ability levels) or ceiling effects (when the test is too easy; Kim et al., 2020).

Material designed for adults may not be engaging to children, and therefore performance may be affected by boredom, fatigue and loss of concentration and attention (Kim et al., 2020). Research has shown that performance improves when the task is enjoyable and interesting (Schukajlow & Krug, 2014) and task persistence and attention has been found to be positively correlated with task enjoyment (Engelmann & Pessoa, 2014; Reeve, 1989).

Standardised assessments can be impersonal and rigid in their administration to minimise confounding factors such as variable levels of assistance from the examiner, social demands, and the involvement of other cognitive processes such as visual processing and attention (McPherson & Burns, 2008). This approach may be stressful for children, and test performance may be impacted by anxiety (Berg et al., 2020; McPherson & Burns, 2008). Tests that examine cognitive functions in clinical settings in isolation may also lack ecological validity as they do not represent real-world application of these demands in everyday life (Wallisch et al., 2018).

Child-friendly tests may include shorter tasks to aid with attention (Howard & Melhuish, 2017) and use child-friendly language to improve understanding and reduce performance being confounded by language skills. Materials can also be made child friendly by making them interactive to improve engagement, such as introducing game-like elements to produce a fun and enjoyable task (McPherson & Burns, 2008). Introducing too many game-like elements may, however, have a negative impact on task reliability, and may reduce construct validity (McPherson &

Burns, 2008). Achieving a balance, therefore, between standardisation and engagement may lead to better task performance that is representative of a child's ability, and may be a more ethical way to assess children (Berg et al., 2020).

1.2.4. Summary of Section

Current tests of CF use a single-trial format or multi-trial format. A benefit of single trial format is that specific skills can be assessed in isolation, however, multi-trial tests are more ecologically valid as they are more representative of day-to-day demands. Tests can also be verbal or visual in nature; visual tests are more culturally fair as linguistic skills and reliance on familiarity with English language can be minimised. Available tests of CF, however, are not culturally fair due to their reliance on familiarity with Western concepts and culture. Such tests were also designed for adults and are therefore not suitable for children. It is therefore possible to use game-like elements to potentially create an assessment of CF specifically for children that is fun and engaging, which is also culturally fair. By doing so, the test will also be a more reliable and valid measure, as it should capture the participants true ability.

1.3. Games As a Source of Engagement

In order to create a game-like assessment, we must understand what a "game" is, what we value about them, and what makes them enjoyable and fun. Nguyen (2017) suggests that the category of "games" is very broad, encompassing computer games, board games, gambling games, sports, child play, live action role-playing games and many more. Each type of game differs in the value it holds as a source of entertainment, and what motivates people to take part. The study of games, however, has mostly focused on sport and computer games, and there is less philosophical study of board games.

Nguyen (2017) suggests that games can be understood as art, sport, or text, and the view adopted changes the criteria used to evaluate them. For example, those who view games as text, may evaluate the value of a game based on its representational content such its narrative and fiction, while those who view them as art may appreciate games based on their authorship and design. Those who view them as sport may view games as an activity, and therefore evaluate them based on aspects such as rules, competition, and fairness. Computer games often involve a rich narrative and detailed graphics and so may lend themselves more to evaluation based on these aspects. Board games, on the other hand, may lend themselves more to evaluation based on rules, competition, and fairness.

Caillois (2001) offers further insight into the question of what a game is, and suggests there are four types of game, including: competitive games, make-believe games, gambling and luck-based games, and games involving disorientation and vertigo (such as that involved in rollercoasters or when children spin in circles). He suggests that some games involve more than one category, for example, the game of poker involves both gambling and competition. He also suggests that the motivation to play games changes across development, for example, a young child may enjoy exploration and impulsive games, whereas adults may enjoy rule-based games.

While the category of “game” is very broad and therefore difficult to define, researchers and philosophers have attempted to identify common elements which can be used to define what makes different games appealing. Formalism defines a game almost exclusively in terms of the inclusion of formal rules to achieve a goal, however, Torres (2018) suggests that this definition is too narrow, and adds that part of the enjoyment of a game is pursuing mastery of a particular skill. Nguyen (2017) argues that another important shared aspect of games is agency; the player pursues the goal of the game while following arbitrary rules, and is motivated to do so by the desire to win. Motivation may be increased by extrinsic rewards and prizes, however, winning in itself holds intrinsic value and reward. Gingerich (2018) additionally suggests that for a game to be a worthwhile activity, it should be challenging in order to provide a sense of achievement when accomplished. Gingerich (2018) also suggests that games hold value as voluntary and unnecessary leisure activities which offer freedom and escapism through a lusory world.

It is important to hold these aspects in mind, therefore, during the development of cognitive assessments involving game-like elements. For an assessment to be fun, it may be important to involve rules and competition in pursuit of the goal. The goals should also be challenging, but achievable, to provide a sense of accomplishment. Additionally, such tests could also involve an element of imagination and fun to improve engagement for children. It is important, however, to balance these game-like aspects with the formality required of an assessment, in order to produce a valid measure CF.

1.4. Literature Review

To situate this research within the current literature, a literature search was completed, and the outcome is reported in the following section.

1.4.1. Methods of Literature Review

Two literature searches were conducted using Scopus, PsychINFO, CINAHL Complete and Academic Search Ultimate databases, for articles concerning the development of tests assessing CF, induction, or abstraction, in primary school children, utilising an approach to increase engagement or enjoyability. The reference lists of relevant articles were then manually searched, and a search was also conducted on Google Scholar to identify related papers.

1.4.2. Search Strategy

The first search strategy involved searching under the major heading (or subject term, or index term, depending on the database), “concept formation” as this allowed exploration of all articles that had been categorised under this subject. To assure no relevant research had been missed, a second search was conducted across all fields using the search terms “concept formation”, “abstraction” and “induction”. For both searches, other search terms remained the same, and included child*, (or paediatric/ pediatric), test (or assessment, evaluation, measure* or instrument) and game (or gamification), whereby the asterisk indicates an abbreviated term.

1.4.3. Inclusion and Exclusion Criteria

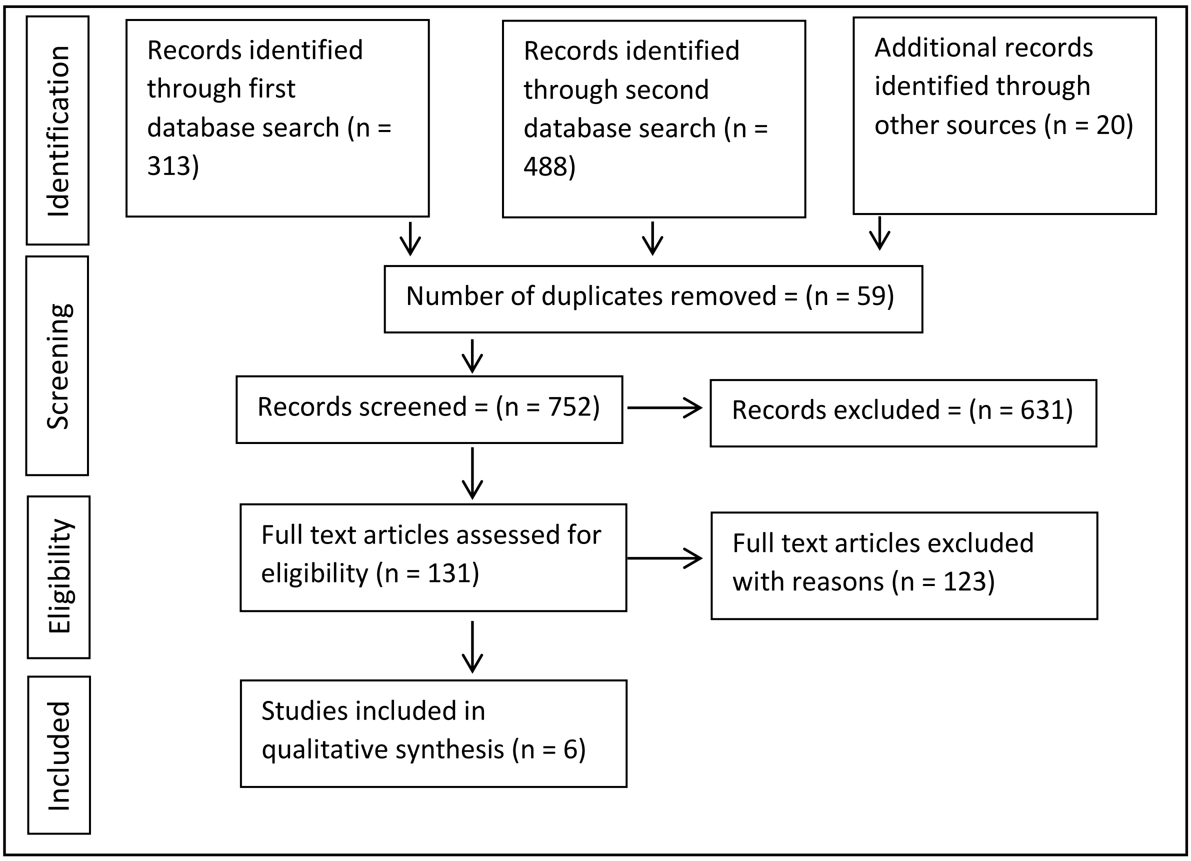
Studies were included if they directly measured abstraction, induction, or CF, included primary school aged participants, and used an approach to improve engagement or enjoyability. Only empirical papers, published in English, and within a peer reviewed journal were included, though the search was not limited to a timeframe.

1.4.4. Search results

A PRISMA Flow Diagram (Figure 1; Moher et al., 2009) was used to illustrate each stage of the literature review and the number of articles included and excluded at each stage. The main search identified 313 articles, 24 of which were duplicates and were therefore excluded from the results. The second search identified a further 488 articles, of which 35 duplicates were removed. Titles and abstracts of the remaining papers were then reviewed using the inclusion criteria described above. Those that appeared to fit the inclusion criteria were included in the final stage which involved reviewing the full paper to determine eligibility. Overall, 6 papers were identified that investigated the assessment of CF in primary school aged children, utilising a game-like approach.

Figure 1.

PRISMA (2009) Flow Diagram of Article Selection Process (Moher et al., 2009).



1.5. Review of Literature

1.5.1. <u>The Tinker Toy Test</u>

Roberts et al. (1995) used the ‘Tinker Toy Task’ to measure problem-solving abilities in children. The task was originally developed by Lezak (1982) to tap into several executive skills including goal formation, planning, initiation, and execution, and Roberts et al (1995) replicated this task for paediatric population. The study recruited 103 children aged 5-12 years from a school in Midwest United States, and 55 primary school age children who had a diagnosis of a mild-to-severe traumatic brain injury (TBI), to test whether the Tinker Toy Task could differentiate these two groups. The TBI group were further subdivided into three groups (‘no impairment’, ’mild-moderate impairment’ ‘severe impairment group’) based on level of cognitive impairment as measured by IQ.

All children completed a neuropsychological assessment battery along with the Tinker Toy task which involved providing participants with a container of 62 Tinker Toy pieces (that included flat pieces, wheels, sticks and connectors), and children were instructed to create something from the pieces. Participants were told they could use as many pieces as they wish and were given up to 10 minutes to complete the task. Following construction, participants were then asked questions about the object they created, such as the function, name, and how they planned and constructed the item. Participants were scored on their answers along with the physical construction produced.

Results showed that the task failed to identify developmental trends in EF in the control group, as scores in the Tinker Toy test did not significantly corelate with age. Additionally, no correlation was found between Tinker Toy task score and intelligence despite a range of IQ scores in the sample. A significant difference in task performance was found, however, between the TBI group and the control group. In the TBI group, Tinker Toy test scores again did not correlate with age, however, scores did significantly correlate with IQ, and the Tinker Toy task was found to be a better predictor of functional impairment than IQ in this group.

The results suggest that the Tinker Toy task could potentially be utilised with children to distinguish problem-solving skills between clinical and nonclinical samples though may not be sensitive enough to identify mild difficulties or developmental trends in non-clinical samples. Furthermore, while this task may provide an overall indication of EF skills, it does not allow for individual skills to be assessed in isolation. While enjoyment and engagement were not measured, the game-like nature of the task could potentially provide a more engaging, and therefore accurate, measure of EFs in children. Additionally, the novel nature of the task suggests this could be a more culturally fair measure of planning and problem solving, however the verbal requirements of the task may disadvantage examinees for whom English is not their first language and those with language difficulties.

1.5.2. <u>The Object Classification Task for Children (OCTC)</u>

Smidts et al., (2004) developed the OCTC as a test of conceptual reasoning in children, based on the Concept Generation Test (Levine et al., 1995), and the Weigl Colour-Form Sorting Test (WCFT; Weigl, 1941). The WCFT consists of a set of stimuli which differ in shape (circle, triangle or square) and colour (blue, red, yellow, or blue). Examinees are asked to sort these stimuli into categories, and those who successfully sort the stimuli once, are asked to sort them again for a second time, thus requiring sorting and set-shifting skills. The WCFT was originally designed for adults and was among the first tests of CF and abstract reasoning using this approach. It has since been used as a fast and easy-to-administer test of executive function impairment, and as it also requires minimal verbal and motor skills, has been used as a test of cognitive impairment with adults who have brain injuries, learning difficulties, and dementia (Hobson et al., 2007). Smidts et al., (2004) developed this assessment to be more engaging to children.

In the OCTC, children were introduced to six plastic toys which differed in some perceptual attributes such as size, colour, or function (e.g., a large red car, a large red plane, a large yellow car, a small red plane, a small yellow car, and a small yellow plane). In the free generation condition, children were asked to make two groups of toys that have something in common and to continue to sort the toys until they could no longer identify common attributes. If a participant struggled, the examiner removed two toys so that the four that remained differed by just two dimensions (e.g., colour and size). For those who continued to struggle, the examiner either sorted the toys for the child and instructed the child to identify the rule they were sorted by (identification condition) or children were given instructions on how they can sort the toys (explicit cueing condition).

The study recruited 84 children aged between 3-7 years-of-age in Melbourne, Australia, and were required to have English as their first language. Results suggested a linear trend between performance and age, where older children performed better. Task performance significantly differed between 3 versus 5-year-olds, and 5 versus 7-year-olds. Children older than 4-years were able to complete the full task, however more than half of 3-and-4-year-olds struggled, and so were given the four-toy condition instead. More than half of these children continued to struggle to sort the toys, and none of these children could sort the four toys using more than one rule. Children older than 4-years who successfully sorted the six toys according to the first identified rule, struggled to sort them again using a second rule, and all children struggled to sort via a third category. The authors suggest this could reflect conceptual or set shifting difficulties or other EFs that are required to successfully complete this task, such as inhibitory control and selective attention.

Overall, the OCTC successfully identified developmental trends in concept development and switching skills, suggesting it could be used as a measure in typically developing children over the age of 4 years, and the use of physical toys may make this test child-friendly and engaging. This test represents a multi-trial test of CF suggesting good ecological validity. The visual nature of the game the use of perceptual categories suggests this game may be more culturally fair than current available CF tests.

1.5.3. The Flexible Item Selection Task (FIST)

In similar aim to the studies previously mentioned, Jacques and Zelazo (2001) developed the Flexible Item Selection Test (FIST) for pre-schoolers as a test of cognitive flexibility and abstraction. The FIST is also a variant of the WCFT (Weigl, 1941), and based on the Visual–Verbal Test (Feldman & Drasgow, 1951). Participants included 197 typically developing children aged 2-5 years who were English speakers.

The task involved 48 laminated cards, each with an image of an item which differed across four dimensions: object, colour, number, or size. Sets of four cards were used in the demonstration and practice trials, and children were asked to identify two cards that matched in some way. Following the practice trials, 12 test trials were presented, each consisting of three cards (one test card and two other cards). The test card (e.g., one medium pink phone) matched one of the other two cards (e.g., one medium purple phone and one medium pink fish) on one relevant dimension (e.g., shape) and matched the other card on another relevant dimension (e.g., colour) while the remaining two irrelevant dimensions remained constant (e.g., number and size). Children were instructed to pick two cards that matched in some way, and then find two cards that match in a different way. The authors suggest that identification of the first pair represents abstraction abilities, whereas selection of a second pair also requires cognitive flexibility.

In the practice trials, age was found to be a significant predictor of performance. Children aged 2-years were excluded from further analysis as 85% of this age group failed at least one practice trial, suggesting the test was too difficult or that they could not understand the task instructions. For the first selection in the test trials, 3-year-olds were found to do significantly worse than 4 and 5-year-olds, though 4 and 5-year-olds performance did not significantly differ. Four-year-old's, however, did significantly worse than 5-year-olds on their second pair selection. The authors suggest that the abstraction component of the second trial is more difficult than the first as this second dimension is likely less obvious than the dimension noticed first.

Overall, findings suggest that this task successfully identified developmental trends in abstraction abilities and cognitive flexibility, though it is unclear whether other factors may have contributed to performance as no other demographics were reported. The article also did not report other objects that were depicted on the cards, making it difficult to assess how culturally fair the stimuli are; however, the task did minimise requirement of verbal responses. While other cognitive demands are required for this task, such as task-switching, the multi-trial format of the task suggests good ecological validity. The task is also child-friendly and uses stimuli that appears affordable to reproduce.

1.5.4. Prototypicality Ratings and Conceptual Development

Alt et al. (2013), developed a novel assessment of CF for children who have language difficulties. The procedure involved asking children to rate how “weird” a series of images on a numerical scale representing ‘normal’ to ‘really weird’. Participants were required to compare the example presented to them, to their benchmark of what they understand a typical example of the item to look like, or their encoded conceptual representation of a category.

The rationale for this method is that individuals with a developed conceptual representation of, for example, a rabbit, would rate images of rabbits that differ in perceptual attributes (short vs long ears, fur colour, fur length, size) as a ‘normal’ example of a rabbit. The authors suggest that children who has a limited conceptual representation of a rabbit, due to having only encoded one type of rabbit (for example, a brown rabbit) under the category of ‘rabbit’ they may rate a white rabbit as ‘weird’.

Their participants included 59 adults aged between 18-39 years, and 59 typically developing children aged 6-9 years. In their first experiment participants were shown images of animals that differed in one of several types of characteristics and were asked to rate how weird the image was. Participants were shown pictures of 20 common animals, including standard versions of these images as well as edited versions that varied in shape, fur/ skin pattern or colour, or facial configuration, and this variation was either slight or significant. Images were shown on a computer screen and participants were asked to press buttons one to nine (normal to really weird).

The results suggested that all participants could distinguish between the three categories (standard, slight variation, and significant variation), but adults were more likely to accurately rate a standard image as normal compared to children, suggesting that children’s conceptual knowledge of animals was more concrete and less accepting of variation compared to adults. In their second study, they compared performance in this task of 17 typically developing children to 17 children with a Specific Learning Impairment (SLI), matched for gender and age. Children with SLI were significantly more likely to rate standard images as weird, though there were no group differences in ratings of images of slight and significant variation, suggesting that the ability to rate standard images as normal may tap into conceptual knowledge ability. They also found that children with stronger verbal and nonverbal abilities, (measured using standardised assessments), were more likely to rate the standard images as normal, and that household socioeconomic status did not predict task performance.

While performance in the test was not compared to standardised measures of CF to ascertain construct validity, results of the study suggest that weirdness ratings have potential as a nonverbal assessment of CF in children as developmental trends were identified. The task was produced specifically for children, though was not game-like in nature and the requirements of a computer programme may make this assessment less accessible as a screening tool. Additionally, while the task required little verbal knowledge, the images used may not be equally as recognisable to children from a variety of cultures and backgrounds.

1.5.5. A novel measure of matching categories for early development

A more recent attempt to measure CF used a tablet-based app with the aim to improve engagement (Condy et al., 2021). The aim of this study was to develop a test that was suitable for use with young children and children with cognitive difficulties, and to minimise language and motor demands by using touch screen to record responses. The assessment was piloted on 15 children with neurodevelopmental delays aged between 2 and 16 years.

The tablet-based assessment involved matching a target item to a response item. The target item was presented at the top of the screen with a set of response items below, and the participant was instructed to drag the target item to the correct response item. Participants were provided with a demonstration and up to three practice trials until a correct response was given, before moving onto the test trials. During test trials, the participants moved up levels as they provided correct responses. As the level increased, the similarity between the target and response items became less obvious and the number of distractors increased. The response items differed by perceptual attributes (e.g., colour, shape, size, or quantity) or semantic attributes (logical relationships such as food items or plants). The task instructions were provided nonverbally, and verbal instructions were minimised.

Results showed that only three of the 15 participants successfully completed the task, while six participants struggled with the motoric demands of the touch screen or struggled to engage with the task, and five had difficulty remembering the task instructions. The authors argue that the study evidenced the feasibility of an interactive CF task via tablet in which language demands were minimised. They suggested further development of the assessment could incorporate additional prompts throughout the trials to minimise memory demands, the use of visual or auditory cues to reinforce correct responses to increase engagement and utilising a different response method.

While developmental trends cannot be ascertained as the task has not yet been piloted on typically developing children, the task is child-friendly and uses technology to improve engagement and reduce language requirements. The study does not describe the images used for the stimuli, however, and the use of semantic relationships may reduce the cultural range of the test. The requirement of a tablet may also reduce affordability and accessibility of the assessment.

1.5.6. The Robots Task

Alderson-Day and McGonigle-Chalmers (2011) developed a test of CF for children with ASD. Previous research has identified categorisation difficulties in children with ASD (Shulman et al., 1995) and in problem-solving tasks such as The 20-Questions task (Minshew et al., 1994)

Alderson-Day and McGonigle-Chalmers (2011) developed two 20-questions style tasks using perceptual and conceptual categories. The authors argue that existing 20-questions style tasks require an understanding of abstract representations and their relational hierarchy, as well as working memory, cognitive flexibility, and language skills. Children with ASD may therefore find categorising based on perceptual categories easier than using conceptual categories, as this requires fewer additional demands (Alderson-Day & McGonigle-Chalmers, 2011).

Fourteen children with a diagnosis of ASD took part in the study, along with 14 age-matched typically developing controls, all of whom scored within normal IQ parameters. The baseline condition involved images that could be organised using conceptual criteria such as tools, living things, and vehicles. The second task (‘Robot Task’) involved novel robot characters that required only perceptual criteria, including colour, shape, head shape, feet, and number of eyes, and each robot had a name written below. These were used to appeal to younger children to increase engagement. The task was set up in a similar style to the children’s game ‘Guess Who?’ where stimuli were arranged on a 6x4 grid with hinged frames. Three trials were used to compare performance, in the ‘initial’ condition, physical elimination of items was permitted to reduce working memory demands and an equal number of items per category was included. In the ‘flexible condition’ items could be removed when they were eliminated, though the category distributions were varied, and in the ‘memory condition’ items could not be removed.

Results showed that ASD participants completed fewer trials in the standard condition than controls, though they completed an equivalent number of trials in the Robot Task. Scores in both participant groups did, however, approach ceiling in the Robot Task, suggesting the task may not be sensitive enough to identify group differences.

Group differences in quality of questions were found. Firstly, in the baseline condition, ASD participants asked more functional questions (questions concerned with the use of the objects) and fewer abstract questions (questions about an abstract feature or taxonomic group) compared with typically developing children. In both tasks the ASD group eliminated significantly fewer items per question and appeared to ask more hypothesis-testing questions which focused on single items (e.g., “is it the bowl?”). Even when constraint-seeking questions were asked (questions which aim to narrow down the options available by focusing on a category, for example, “is it a living thing?”), the ASD group still eliminated fewer items per questions, and repeated more questions.

When the item elimination was not permitted, so the whole array was always visible, the group differences increased: ASD participants asked significantly more questions, fewer abstract questions, and more functional questions compared to controls. This suggests that ASD participants struggled with the additional working memory demands, which impacted their performance.

In the flexible condition, where category distribution varied, question quality in the control group improved somewhat, though this was not a large change. Otherwise, no significant group differences were found, suggesting that category proportions did not impact task performance for either group. Lastly, verbal abilities were found to be a significant predictor of score for the ASD group, whereas only age was a predictor in the control group.

Overall, the tasks identified developmental trends in the typically developing group, and children with ASD could be identified based on qualitative differences in question quality, suggesting this to be a promising measure of conceptual skills. A ceiling effect was reached in the Robot Task, however, suggesting the game may have been too easy, and might be improved by increasing the number of perceptual criteria used.

The Robot Task appears to be a better test of CF than the standard 20-questions style task due to the use of perceptual categories, and the game-like nature of the task is likely to be appealing to children. Perceptual categories appear to make the task more accessible for children with ASD and the use of novel images, along with limited language demands, is also likely to be less culturally specific.

1.5.7. Summary of Literature Review

The literature search identified six studies that were focused on developing a child-friendly assessment of CF. All tasks used a method to improve engagement to make their assessments child friendly, such as using physical toys, child-friendly pictures, use of technology, or a game-like structure. None of these studies, however, measured enjoyability nor reported acceptability feedback from participants.
Four of the six studies identified developmental trends in task performance which suggests the task could be used to assess EF difficulties, while the other two tasks failed to identify developmental trends in a nonclinical sample (Tinker Toy Task; Roberts et al., 1995) or a ceiling effect was found (Robot Task; Alderson-Day &

McGonigle-Chalmers, 2011), suggesting these tasks may not be sensitive enough for children across the age range. None of these tasks, however, compared performance to established measures of CF to assess construct validity.
None of the studies identified for this literature review mentioned cultural fairness as a factor considered in the development of their assessments, however, those tasks which used novel visual stimuli, perceptual categories, and limited verbal requirements such as the OCTC (Smidts et al., 2004) and the Robot Task (Alderson-Day & McGonigle-Chalmers, 2011) could be argued as more culturally fair than current measures of CF.

1.6. The Alien Game

Pavitt (2017) designed The Alien Game (Appendix A) as a new measure of CF in children using novel images of aliens, and piloted this test on 18 children aged 7-11 years. This task was based on Alderson-Day & McGonigle-Chalmer's (2011) Robot Task, and participants were instructed to identify which alien the examiner was thinking of by using yes or no questions.

Images of aliens were used with the aim to be more culturally fair than the stimuli used in the propriety game "Guess Who?. While the concept of cartoon illustrations of aliens is still rooted in Western culture, the rationale for using aliens is that they can be designed by the researcher to be novel to all participants to avoid familiarity with the stimuli advantaging some participants. The aliens also differed by perceptual categories (such as colour, shape, and number of limbs) which are less culturally specific than the attributes used in the propriety game "Guess Who?.

Similar to The Robot Task, the Alien Game used the format of the children's game 'Guess Who?' to improve engagement, by arranging the images on a six-by-four grid on a plastic apparatus so that items could be physically eliminated. Whereas the stimuli in The Robots Game differed by five characteristics, the stimuli in The Alien Game differed by 15 attributes to increase the difficulty of the game, and some of the characteristics had two options (e.g., no tail or one tail) or multiple options (e.g., no legs, two legs or three legs). In Pavitt's (2017) study, the characteristics were unevenly distributed to add a chance element, to make the task more game-like, and each image had a made-up name below (for example, Parf, Anbe, and Utti).

The study found that typically developing children adopted a similar strategy in the task and that a normative pattern of responses could be established, representing typical strategies and variation in CF skills. This study also measured task enjoyability, and found that children rated the game as more enjoyable than established measures of CF. Following this pilot study, Pavitt (2017) suggested several improvements and future study directions. Firstly, Pavitt's (2017) study tested a small sample of children in which only two participants had English as a second language. This sample was therefore not representative of the London population and could not measure whether factors such as first language, cultural background / identify, and SES affected task performance. It was also suggested that task score could be correlated with other multi-trial tests of CF to establish concurrent and predictive validity.

1.7. Current Study

The current study aims to further develop The Alien Game (Appendix B) based on the above suggestions, and by developing the physical stimuli and a scoring method, and to conduct a second pilot study.

1.7.1. Research Aims

The aims of this study are:

- To refine the materials used in the first pilot study by evening out the distribution of the alien characteristics so a single question cannot remove more than half the aliens, to remove the element of chance and produce a more structured test.
- Remove the alien names under the image to further reduce language requirements.
- To replace the plastic apparatus and instead print the images on an A3-size grid. This will improve accessibility and affordability of the test and remove the ability to physically eliminate items. This may add working memory

demands, though will make the test more stringent and improve ecological validity as it will be more representative of day-to-day demands. A second grid will be printed and cut into individual cards to allow for physical elimination of non-targets if children struggle with the working memory demands of the first format.

- To test this game on a larger sample, focusing on a specific age range (8-11 years) to investigate whether ceiling effects are reached in this age group.
- To pilot on a classroom of children from a London-based school with the aim of recruiting a more culturally diverse sample to gather richer data on the feasibility of this game as a culturally fair test of concept formation.
- Compare performance on The Alien Game to established measures of concept formation to test concurrent validity.

1.7.2. Research Questions

The research questions for this study are:

1. Can a culturally fair test of concept formation be derived from the Alien Game, for children aged 8-11 years?
2. Will the test format be acceptable and engaging for children?
3. To what extent will Alien Game scores correlate with established measures of concept formation, such as WISC-IV Matrix Reasoning and Similarities (Wechsler, 2003)?
4. To what extent will the Alien Game scores correlate with objective reports of executive functions in children, such as in the CHEXI (Thorell & Nyberg, 2008)?

2. EPISTEMOLOGY & METHODOLOGY

The following section will begin by exploring the epistemological position adopted by the researcher and the methodology of the study. This will include a description of how the Alien Game was developed, followed by a description of how data was collected, and how this data was used to develop a scoring procedure for the Alien Game. Finally, this section will set out participant demographics and any group differences in the sample which could impact the interpretation of the results.

2.1. Epistemology

Epistemology is concerned with the theory of knowledge and how it is acquired (Audi, 2011). The epistemological position adopted by a researcher is based on their value set and world view, and comes with a set of assumptions (Willig, 2012). Such assumptions will influence the research process including how data is collected, analysed, and interpreted. It is therefore important to state the epistemological position that is taken, so the accompanying assumptions can be scrutinised and potential limitations of the research can be identified (Darlaston-Jones, 2007).

There are several epistemological perspectives, including realism, constructionism, and phenomenology (Willig, 2012). Phenomenology is concerned with subjective experience rather than an objective truth, while constructionism assumes that what is “true” changes over time depending on the sociocultural context. Realism, on the other hand, assumes the existence of “true” knowledge, and that all entities, natural and social, are measurable objective phenomena. Critical realism can be considered as a middle ground between realism and constructionism, and assumes the existence of real knowledge that can be measured, while acknowledging that reality is constructed by the current social-cultural place and time, and is therefore changeable in nature (Bhaskar, 2010; Gorski, 2013). The underlying assumptions of an approach will determine how data is collected, analysed, and interpreted, for example, a realist position may use quantitative approaches to measure objective data, whereas constructionist approaches may use qualitative approaches to examine subjective meaning.

The epistemological stance adopted by the researchers involved in this study is critical realism. By developing an assessment measure, we are assuming that cognitive abilities, such as concept formation (CF), are real measurable entities that can be objectively assessed. We are, however, critical in our approach and acknowledge that our understanding of this capacity is situated within our current socio-cultural context, and thus the understanding of 'concept formation' may change over time and be understood differently in other cultures. This research and underlying assumptions should therefore be situated within the UK context in which the research was carried out, and caution should be applied if extending the findings to other contexts. As this study aims to measure what is "real", a quantitative experimental approach was adopted, using correlational research design to analyse the data.

2.2. Design

This study aimed to develop the Alien Game, and to establish a scoring protocol. Following development of the Alien Game, it was tested on a sample of children, and a scoring procedure was developed. Exploratory data analysis was used to interpret this data, and then to test the feasibility of this assessment, and scoring procedure, as a child-friendly test of CF. A cross-sectional research design was then used to address the relationships between each scoring method, to establish whether they are testing different skills; and between performance in The Alien Game and established CF measures to look at concurrent and predictive validity. Observational and qualitative data was also gathered to assess changes made to the game, and feedback was collected in the form of Likert-scale responses to determine task enjoyment.

2.3. Materials

The materials used include:

- The Alien Game testing pack which included:
  - Participant information sheet (Appendix C).
  - Participant consent form (Appendix D).
  - Participant debrief form Appendix E)

    - Demographics record form (Appendix F).
    - Alien Game response record form trials A-D (Appendix G).
    - Likert/visual analogue scale to rate task enjoyability (Appendix H).
    - WISC-IV Matrix Reasoning and Similarities subtests (Wechsler, 2003) and scoring sheets as measures of concurrent validity.
    - Teacher-rated CHEXI (Thorell & Nyberg, 2008) as a measure of predictive validity (Appendix I).
- Research advertisement (Appendix J) .
- Parental information sheet (Appendix K).
- Parents opt-out consent form (Appendix L).
- Parent debrief form (Appendix M)
- Loco parentis consent form (Appendix N).
- Pen and paper.
- Standardised instruction sheet (Appendix O).
- The Alien Game (Appendix B).

This study aimed to improve and refine the materials and procedure used by Pavitt (2017) in the original pilot study of this assessment. Before data collection is described, the next section will outline how the Alien Game was developed.

2.4. Development of The Alien Game

Pavitt (2017) designed The Alien Game as a game-like measure of CF for children, to improve engagement (Gioia, 2015). The assessment is based on the traditional 20-Questions task which involves guessing the examiners target from an array of items by asking questions. The Alien Game used visual images of aliens which differ on several characteristics: for example, their colour or shape, and the number of eyes, antennae, legs, or arms. In the original version of the Alien Game, the pictures were placed on a plastic apparatus similar to the children's game 'Guess Who?' to make the test more game-like and to allow for targets to be physically removed from sight after elimination. It was found that children approached the task in a consistent manner, and reported that they enjoyed the game, though a few areas were identified which required development. This included the assessment stimuli, the scoring procedure, as well as exploring the validity of the assessment. In this study, I, along with two other researchers, aimed to address these areas. We met together as a group to collaboratively re-design the alien illustrations using computer software, and each of us contributed equally to this. Myself and one other trainee then used these illustrations to continue with the present study. We independently tested the game with participants of different age groups; I collected data from children aged 10-11, while the other researcher collected data from children aged 8-9. We then merged our data and met together to collaboratively create a scoring procedure. We then independently scored the combined data and ran the analysis, and then met together to compare results to ensure accuracy. This next section will describe the development of the stimuli, and the rationale for the decisions we made.

2.4.1. Illustrations

Pavitt (2017) created 24 custom-drawn aliens (Appendix A), so that all stimuli were novel to each participant to eliminate practice effects due to familiarity with proprietary game formats. In Pavitt's (2017) study, however, it was noticed that some attributes of these aliens were difficult to distinguish and identify, and so the materials were refined in this current study to address this (Appendix B). For example, the alien's wings in the original version were somewhat unclear, and could be confused with ears, and so these were developed by the researchers in this study (see Figure 2). Two attributes were also removed (skin texture and eye colour) in this study, so that the aliens differed by a total of 13 attributes rather than 15, with the aim to make those which remain visually clearer (see Figure 3). Similar to Pavitt (2017), the colours of the aliens (yellow, blue, and green) were chosen so that any participants who were colour-blind would be able to differentiate the colours.

Figure 2.

Example of Illustration Development: Ears and Wings

Note: Original illustration (left) modified (right) to make attributes more distinguishable, including ears and wings.

Figure 3.

Example of Illustration Development: Skin Texture and Eye Colour

Note: Original illustrations (left) developed (right) to remove variation in skin texture and eye colour to make remaining attributes easier to identify.

Six of the 13 attributes via which the aliens differed included three options:

- colour (blue, grey, or yellow)
- shape (circle, square, or triangle)
- eyes (one, two, or three)
- antenna (zero, one, or two)
- legs (zero, two, or four)
- nose (none, a small nose, or a long nose).

The remaining seven attributes differed by only two options and were either present or not present: horns, tail, ears, wings, teeth, arms, and eyebrows. Additionally, in Pavitt's (2017) game, an uneven number of aliens were assigned to each attribute, to introduce an element of chance, with the aim to increase the game-like nature of the assessment to prioritise engagement and enjoyability. In this study, however, the attributes were evened-out so that each category contained the same number of aliens (for example, half of the aliens had wings and half did not, and there were the same number of aliens that were blue, yellow, and grey). The rationale for this was to prioritise producing a fair assessment of CF skills, and to test whether this impacted enjoyability. Another adaptation was to remove the names of the aliens used in the original game, and use numbers instead, to reduce language demands with the aim to make the game more language fair.

2.4.2. Game Apparatus

In the original game, Pavitt (2017) arranged 24 pictures of aliens in four rows of six on a plastic board with hinged frames, similar to the setup of "Guess Who?" (Alderson-Day & McGonigle-Chalmers, 2011). This also allowed for physical elimination of the aliens to reduce working memory demands. In the current study, a plastic apparatus was not used and instead the aliens were printed and laminated on A3 paper to reduce cost and to make the assessment affordable and accessible to reproduce. In this study, two versions of the game were created; in one the aliens were printed on A3 paper and arranged in a six-by-four grid so physical elimination of the aliens was prevented. In the other version, this grid was cut into 24 individual cards which were arranged in four rows of six, on a table in front of the participant, and could be turned upside down to physically eliminate non-targets. These two versions were made to tailor the game to the developmental level of the child, so the grid (which requires working memory and is therefore harder) could be used for older children, while the cards (which reduce working memory demands and is therefore easier) could be used for younger children.

2.4.3. Game Procedure

This study involved four trials in keeping with Pavitt's (2017) Alien Game. In the original version of the game, a trial was discontinued after 10 questions; whereas a discontinue rule was not set in this study, in order to explore the range of questions required to reach the correct target. Similar to Pavitt (2017), simple instructions were developed to be suitable for the participant's developmental level (provided in Appendix O).

2.4.4. Summary of Development of the Alien Game

In summary, the Alien Game was developed in several ways following Pavitt's (2017) pilot study. Firstly, the illustrations used by Pavitt (2017) were developed to make the alien's attributes clearer and easier to identify. This also involved removing some attributes such as skin texture and eye colour, to make the remaining attributes clearer. While Pavitt (2017) prioritised the game-like nature of the assessment by including an element of chance, we removed features to prioritise creating a fair assessment. This included evening out the alien's attributes, to reduce the element of chance in correctly identifying the target. Additionally, the plastic "Guess Who?" apparatus was not used in this study: the aliens were printed and laminated on A3 paper and cards. Lastly, we removed the discontinue criteria of 10 questions to explore how many questions children might require to reach the correct target.

In order to assess the feasibility of the Alien Game as an assessment of CF, the game was tested on a sample of typically developing children, and the data was used to explore possible scoring methods. The following sections will outline the data collection procedure.

2.5. Inclusion & Exclusion Criteria

The author of this thesis recruited and collected data from children aged 10-11 years from a mainstream school in London, while another researcher conducted this same study with children aged 8-9 years from the same school. Each researcher focused on a specific age range as children's cognitive capabilities develop rapidly during childhood, and so it was hypothesised that the approach to administering such a task may differ from one year group to the next, and adaptions may need to be made for younger children compared with older children. The researchers felt a focused, and in-depth look at each age group would allow for a cognitive assessment to be created that would aid with engagement and consider the child's developmental age, as to prevent a ceiling of floor effect in test scores. The upper age limit of 11 years of age was chosen as children's executive functions begin to mature at 12-14 years of age (Lord-Maes & Obrzut, 1996) and children begin to be better able to tolerate assessments such as those which exist for adult populations but have children's norms. The importance of developing an assessment that is enjoyable and engaging is therefore most important for children younger than 11.

This study aimed to capture typical variation in CF abilities that exist in children attending mainstream schools. We therefore aimed to accommodate all needs as much as possible, to not exclude any participants. The only reason a student was excluded from the study was if their ability to consent to participating was unclear due to limited understanding of written and verbal information in English.

2.6. Ethics

Ethical approval was obtained from the University of East London's Research Ethics Committee (Appendix R). Information sheets and opt-out consent forms were created for parents (Appendices K and L), and a child-friendly version of each was also created to meet the developmental age of the child, including more colour and images (Appendices C and D).

Participants were recruited from one school in London. The researchers contacted various schools within Greater London via email, with details of the study and the study advertisement (Appendix J) and were asked to reply to the email if they would like to take part. One school replied to this email with interest to participate. The head teacher at this school chose an opt-out procedure, so parents were sent an information sheet with details of the study and were advised to return the signed opt-out form by a given date if they did not wish their child to participate. Otherwise, the teacher acted in *loco parentis* for those children who were not opted-out by their parents.

The researcher met with each child individually and read through the child-friendly information sheet and consent form with them, and allowed the child the opportunity to ask questions, to ensure they understood what was being asked of them, or to take some time to consider participating if they were unsure. Those who agreed to participate were asked to sign a child-friendly consent form and were told they could stop participating at any point without giving a reason why.

Parents, the head teacher, and the students, were also informed that they had until the end of January 2023 to withdraw data if they changed their mind, and could do so by contacting the researcher, and were fully informed about what participation in the study would involve. The study did not involve any risks or hazards and no participants appeared distressed at any time during the study.

2.7. Confidentiality

Consent forms, which included the child's name and date of birth, were kept separately from all other information in a locked cupboard within the deputy head-teachers office. These were also scanned, and a copy saved to the researcher's UEL secure IT account. Each child was assigned a random identification number, which was recorded on a password-protected MS Excel Spreadsheet and saved in a separate folder on the researcher's UEL secure electronic cloud, so each child's data could be identified if they wished to withdraw from the study. After the end of January 2023, when results had been analysed, the folder containing consent forms and identification numbers was deleted, and the school were instructed to securely destroy the consent forms by shredding in accordance with UEL guidelines. All other records (e.g., test forms, Alien Game record form, and demographics) included the participant ID code in place of the child's name. The data from these forms were stored on a MS Excel (Microsoft Corporation, 2018) and SPSS database (IBM Corp, 2021) and saved on the researcher's UEL secure electronic cloud. This database did not include any personally identifiable data.

2.8. Procedure

Testing took part in a quiet room at the primary school. Children were called one by one from their class and were given a child-friendly information sheet which was read with them. They were then asked if they would like to take part in the study and were told they could say no at any point without giving a reason. Some children declined, and were accompanied back to their classroom, others asked for some time to consider, and others agreed straight away. Those who agreed were asked to read through the consent form and asked to tick 'yes' or 'no' to each question. During this process it became clear that two children, who had very limited English language as they had recently migrated to the UK, did not fully understand what was being asked of them and appeared to tick the boxes on the consent form in a random way. These children were deemed unable to consent to participate in the research and so were accompanied back to their classroom. Those who understood the task, and were deemed able to consent, were introduced to the Alien Game. Each child was read a standardized set of instructions (given in appendix O) and were given the opportunity to ask any questions they had.

2.9. Data Collection and Analysis

The children's teacher was given a CHEXI questionnaire (Thorell & Nyberg, 2008; see Appendix I) form to complete which is a rating instrument for measuring executive functions, including working memory and inhibition. Responses to the 24 questions included in the CHEXI use a 5-point Likert scale, from "definitely not true" to "definitely true".

All participants completed four trials of the Alien Game, and the order of the target alien across the four trials was kept consistent (targets 8, 16, 17, then 24). During administration of the Alien Game, if the participant asked a question regarding the number of the alien (for example, "is the alien above number 12?"), they were told that the number next to the alien was not relevant to the game. If they continued to ask about the number's following this prompt, this was recorded as an ineffective question (see below). If a participant asked the wrong type of question (for example, one that did not allow a yes or no response), this was also recorded as an ineffective question, and they were reminded of the task rules which stated that the researcher could only answer 'yes' or 'no'. If a participant asked an ambiguous question (for example, "does your alien have limbs?") they were asked to clarify the question and both responses were recorded. Following completion of the Alien Game, participants were provided with a visual analogue scale (Appendix H) and were asked to rate the enjoyability of the game from one (awful) to five (fantastic). The participants game enjoyability rating was recorded, as well as each question the child asked.

WISC-IV Similarities and Matrix Reasoning subtests were then administered following the Satz-Mogel method (Satz & Mogel, 1962). These are established single-trial measures of CF and data was collected to test for concurrent validity. All children completed these subtests.

Demographic information was collected from each participant, including their date of birth, gender identification (referred to as 'sex' from here onwards as every child identified with the gender assigned to them at birth) and primary language spoken at home, as well as other languages spoken, sensory or motor needs, nation of family origin, and father's occupation. Fathers' occupation was collected rather than mothers' as research shows that fathers occupation has a more significant impact on family health, SES, and educational outcomes (Pinilla et al., 2017). Following data collection, parental occupation was graded via the socio-economic classification criteria outlined by the NRS (National Readership Survey, 2006).

Each child completed the WISC-IV Similarities and Matrix Reasoning subtests (Wechsler, 2003), and their raw scores were converted into scaled scores (SS) for analysis. The teacher-rated CHEXI (Thorell & Nyberg, 2008) was scored and converted into two domains (working memory and planning) based on the two factors identified by Catale et al. (2015). While normative data has not yet been established for the CHEXI, Catale et al (2015) collected data from a sample of 242 typically developing children, and scaled scores were developed from this data. These transformations were used in this study to provide an idea of how the children in this sample scored in relation to Catale's (2015) sample.

2.10. Scoring Procedure

To create a scoring procedure for the Alien Game, we considered how to best capture variability in task performance, and therefore the best measure of CF skills and problem-solving abilities. We considered how to best measure the strategy children adopted by considering which type of question was the most effective to ask at each point in the game (e.g., hypothesis-testing (HT) questions versus constraint seeking (CS) questions). CS questions are those based on categories, with the aim to remove a large number of options in order to constrain the search (for example, is your alien blue?); the ability to categories in this way requires conceptual skills. HS questions, on the other hand, are those in which a question is asked about a single item (for example, is it number 17?; Alderson-Day & McGonigle-Chalmers, 2011).

We also considered how conceptual and problem-solving skills were employed throughout the game (for example, how many questions were required to reach the target, how many targets a question eliminated, and the frequency and type of errors made). User-friendliness of the scoring criteria was also prioritised, and so we considered the pros and cons of using complex detailed scoring methods (which may be more accurate though may be difficult to calculate) versus broader scoring methods which may be easier and simpler to calculate by hand. We also reviewed Pavitt's (2017) scoring method, as well as the scoring criteria used in the D-KEFS 20 Questions task (Delis et al., 2001), for suggestions. The next section will describe the scoring procedure adopted, and the rationale for decisions made.

2.10.1. Total Questions

If the best possible strategy is adopted, the target alien should be identified in the fewest possible number of questions. We therefore introduced a measure of the total number of questions asked, named 'Total Questions' (TQ) as a broad overview of performance in each trial. We considered deriving the mean TQ over four trials to create a score representing average performance, however, this greatly constrained scores which were harder to interpret. Instead, we felt that adding questions asked across the four trials would be the most user-friendly approach. This measure, however, does not tell us much about the strategy adopted, nor whether the correct answer was reached out of luck rather than due to exercise of CF skills. Further measures were therefore used to address strategy more closely.

2.10.2. Weighted Achievement Score (WA)

Some participants may guess the correct answer by chance in few questions, and so the weighted achievement (WA) score can be calculated to take this into account. Using this method, points are provided on a u-curve: fewer points are rewarded for reaching the answer with too few or too many questions (see Table 1). This method was used in the DKEFS 20 Questions subtest, and the ideal number of questions to identify target had been derived from an adult normative sample. This point system may therefore need to be updated as more participant data is collected.

Table 1

Weighted Achievement Score Transformation

TQ	WA
1-2	0
3	3
4-5	5
6	4
7	3
8	2
9-10	1
10+	0

Note. TQ = Total Questions; WA = Weighted Achievement Score.

2.10.3. Abstraction Score (Ab)

The Abstraction Score (Ab) represents the level of abstract thinking behind a question asked and is calculated by looking at how many non-target items would be removed by a single question regardless of the answer. We considered several methods to calculate the Ab which are outlined below.

2.10.3.1. *Minimum, maximum of difference*: We considered whether to calculate the Ab score by looking at the most possible items which could be removed by a question, or the fewest, or the difference between the two. As part of this, we thought about which type of question would represent the most efficient strategy, or best conceptual skills:

- *High-risk question*, such as one which might eliminate a large number of targets but risked eliminating only a few, for example, the hypothesis-testing question "is your alien number 15?" could remove 96% of targets on the first question if the answer is 'yes', though risks only removing 4% if the answer is 'no'.

- *Medium-risk questions*, for example the constraint seeking question “is it yellow?” could remove 66% of the targets on the first question if the answer is ‘yes’, or 33% if the answer is ‘no’.
- *Low-risk questions*, for example, “does it have wings?” removes 50% of the targets on the first question regardless of the answer.

We concluded that a high-risk question (e.g., “is your alien number 15?”) would be the worse strategy to use, as such questions did not reflect conceptual skills, since the ability to abstract categories in the array of pictures is not required. We also therefore ruled out using the maximum number of targets which could be eliminated as the abstraction score. We felt that both medium and low risk questions represented good strategy, as both use constraint seeking questions, so we considered using minimum for Ab or the difference between minimum and maximum as the Ab.

As part of this decision-making process, we considered the user-friendliness of the scoring criteria for a future test user and felt that calculating the difference score would be time-consuming and difficult to do by hand. Taking everything into consideration, therefore, we decided to use the lowest number of items which can be removed by a single question as the abstraction score. A large Ab score, therefore, would represent the best strategy.

2.10.3.2. *Proportion vs actual Ab:* We explored calculating Ab using the proportion of targets that could be removed by a question; that is by dividing the minimum number of targets a question might remove by the number of targets remaining. Alternatively, we also considered calculating Ab via the actual number of targets a question could remove. The rationale for using the proportion value rather than the actual value was that as the pool of potential targets reduce over the course of the trial, the actual number of targets that would be removed will also reduce. Therefore, actual (absolute) Ab will inevitably reduce over the course of the trial, even if strategy does not change, whereas the proportion will not.

On the other hand, proportion values are more complicated to calculate compared than the actual value. When we analysed the data using proportion values, we found that the range of scores were very small which would make interpretation difficult and reduces how user-friendly the assessment is. Additionally, we felt that calculation and interpretation of this measure may require software to calculate which may also increase cost. Taking everything into consideration, we decided to use the actual value to make the assessment more user-friendly.

2.10.3.3. *Calculating Ab using every question asked versus first 3 questions:* We also explored calculating the total Ab score for every question asked in each trial or for only the first three. In the end, participants will remove the same number of non-targets when they identify the correct alien, so adding the actual Ab across all questions asked would not create a meaningful value. We therefore decided to calculate Ab based on the first 3 questions only.

2.10.3.4. *Calculating overall Ab score by totalling or averaging across trials:* We considered creating an average Ab score so this value could be compared to individual scores and trials; however, we felt that this would again be difficult to calculate. Difficult calculations may introduce errors in scoring and may reduce how user-friendly the assessment is. We therefore decided to sum the Ab score over the first 3 questions of each trial, and total these together across the four trials, to create an overall Ab score. We felt that this balanced the need for an accurate measure with user-friendliness.

2.10.3.5. *Summary of Ab score:* To balance accuracy of the measure and user-friendliness, Ab was calculated by adding together the minimum number of targets that can be removed by the first three questions of each trial.

2.10.4. Initial Abstraction Score (IA)

The initial abstraction score (IA) was used in the D-KEFS 20 Questions task as a measure of abstract thinking behind the first question of each trial. It is suggested that in a new trial, the participant is faced with the task of identifying the correct target out of 24 possibilities, and so it is important for the first question to be as efficient as possible. The most efficient first question, representing the highest level of abstract thinking, will therefore remove a large number of possible targets. As this is calculated using only the first question, this scoring method has potential to be easier to score than Ab and therefore more user-friendly. In the Alien Game, we calculated IA score by adding together the actual number of items that might be eliminated via the first question of each trial, similar to the calculation for Ab.

2.10.5. Ineffective or Unallowed Questions (IU)

Another way to assess performance is to measure ineffective and unallowed questions. We originally called this measure "errors", however, we felt that 'ineffective or unallowed questions (IU)' was a more accurate descriptor for the question types we included. We identified these as questions which:

- do not remove any targets
- are repeated questions
- are questions which do not allow for a 'yes' or 'no' response and thus violate
- are a wrong guess (HT questions the game rules)

In the original version, Pavitt (2017) counted "compound-questions' as error questions, which are those which combine two or more questions (for example, is the alien yellow and does it have wings?"). We did not consider these question types as ineffective in this study, as we felt that while this question posed challenges for the examiner (as this question could not always be answered 'yes' or 'no'), we felt that examinees were still using conceptual skills to formulate this type of constraint-seeking question. We also did not specify that this was not allowed in the task instructions, so felt it was not fair to reduce a participants score based on this. During the game if this question type was asked, participants were prompted to break the question down. The rationale for the question types deemed ineffective are as follows:

- *Questions which do not remove any non-targets*: We hypothesized that questions which do not remove any options occurred when participants struggled to hold in mind previous questions asked, or the targets that previous questions removed.
- *Repeated questions (RQ):* Similar to above, we hypothesized that participants may repeat questions if they fail to hold in mind previous questions asked.

- *Questions that are not a yes/no format*: These questions violate the rule of a yes/no question and suggests that the participant did not hold the task rules in mind.
- *Hypothesis-testing (HT) questions:* While the target item may be identified out of luck, using HT questions represents poorer problem-solving abilities, so we considered HT questions as ineffective, with the assumption that fewer of these questions reflects better conceptual skills.
- *Total IU questions*: The question types described above were recorded individually. However, RQ, and questions that were not in yes/no format, did not prove to be useful measures independently as these question types rarely occurred (see Table 2). We therefore decided to add these question types together to create a single measure. Some questions could be categorised as two types of ineffective questions (e.g., a HT or RQ may also fail to remove any targets), though we decided to count these only once as to not artificially inflate this score.

Table 2

Total Number of Ineffective or Unallowed Questions

	Frequency	Percentage
Total Hypothesis-testing questions	124	15.12%
Total questions which do not remove any targets	630	76.83%
Total RQ	44	5.37%
Total questions which are not in a yes/no format	22	2.68%

Note. Numbers represent total number of these question types made by all participants across the four trials.

2.10.5.1. *Weighted Ineffective or Unallowed Questions Score (WIU):* Similar to the weighted achievement score, we calculated a weighted score for ineffective or unallowed questions where more points are given to those who ask fewer (see Table 3). This was so that an overall game performance measure could be created based on WA score (calculated using TQ's and provides insight into overall performance) and IU score (which provides insight into strategy and quality of questions).

Table 3

Weighted Ineffective Questions (WIU) scoring.

Total IU	WIU score
0	4
2	3
3	2
4	1
5+	0

Note. IU = ineffective or unallowed questions; WIU = weighted ineffective or unallowed questions

As a raw score, a higher IU score denotes worse performance, whereas a higher WA score suggests better performance. WIU was therefore introduced to transform IU on a graded system, so a higher score is suggestive of better performance in a similar way to WA. WIU and WA could then be added together to produce the Alien Game score.

2.10.5.2. *Summary of IU questions:* ineffective questions were calculated by adding together the total number of questions which do not remove any options, RQ, questions not in yes/no format, and wrong guesses (HT questions). A weighted score was then derived by transforming these on a graded point system.

2.10.6. Alien Game Score (AG)

As an overall indicator of game performance, we created an AG score which is the sum of the WA and WIU as previously described. The rationale for this additional measure was that the WA is based only on the number of questions asked, and while this addresses the possibility of a lucky guess, it does not consider strategy.

Ineffective questions, however, may provide a good measure of the ability to follow the game rules, and hold in mind the answers to previous questions asked, and thus provides a good measure of capacity and as a measure of strategy. We therefore hypothesised that a measure incorporating both would be a fuller measure of game performance (than each measure alone) and would be easy to calculate.

2.10.7. Summary of Scoring Procedure

- **Total questions (TQ):** calculated by adding together the total number of questions required to reach the correct target across the four trials.
- **Weighted achievement score (WA):** calculated by scoring TQ's on a u-curve (see Table 1) so fewer points are rewarded for reaching the answer with too few or too many questions.
- **Initial abstraction score (IA):** calculated by adding together the minimum number of targets that could be removed by the first question of each trial.
- **Abstraction score (Ab):** calculated by adding together the minimum number of targets that could be removed by the first three questions of each trial.
- **Ineffective or unallowed questions score (IU):** this measure was calculated by adding together the total number of questions which do not remove any options, RQ, questions which are not in a yes/no format, and wrong guesses (HT questions)
- **Weighted ineffective or unallowed questions score (WIU):** This was calculated by scoring ineffective questions on a graded system (see Table 3).
- **Alien Game Score (AG):** This was calculated by adding together the WA and WIU's score.

2.11. Alien Game Evaluation

2.11.1. Acceptability

The acceptability of the Alien Game was measured by asking participants how enjoyable they found the game by rating on a visual analogue Likert-scale from 0-4 (awful to fantastic). This data was collected for each participant and the frequency of each score was explored.

2.11.2. <u>Evaluation of Attributes</u>

To assess the development of the illustrations and whether the characteristics of the aliens are identifiable and distinguishable, we explored the questions asked concerning each attribute.

2.12. Participants

The author of this study recruited and tested 17 children aged 10-11, while another researcher recruited and tested 17 children aged 8-9. We combined data for these two age groups (8-9 and10-11) to create a larger sample, and to explore whether there are any age-related differences in performance.

In total, 34 children took part in this study. One participant was removed from analysis as their data showed as an outlier. Closer exploration of their strategy revealed that their questions were often ambiguous, and it was unclear whether their strategy was poor, or if the researcher had misunderstood their questions during administration of the Alien Game. The total number of participants used within the analysis was therefore 33.

All participants were recruited from the same primary school. Participants age ranged between 100.11 and 132.63 months (M = 117.66, SD = 11.73). In total, 16 males and 17 females participated. As this study aimed to create a culturally fair assessment which does not advantage children from a Western European culture, participant's ethnicity was recorded which is summarised in Table 4. This data shows that a diverse participant group took part in this study; 76% of whom reported a family background categorised as Western European, while 24% of the sample reported growing up in an Eastern-European or non-European country. In comparison to the ethnic composition of London (Office for National Statistics, 2022; see Figure 4), the sample had slightly fewer Asian, White British and Other participants, and more Black, Mixed and White Other participants (consisting mostly of Western European ethnicities).

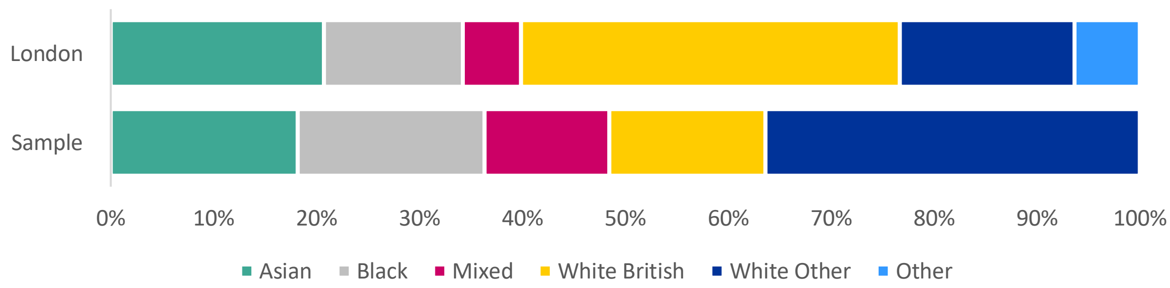
Table 4

Participant Ethnicities

Group	Ethnicity	Frequency	Percentage
Western European	White British Other	9	27.27
	White British	5	15.15
	Black British	4	12.12
	British White & Black Caribbean	2	6.06
	Polish	2	6.06
	British - White & Asian	1	3.03
	British Mixed Other	1	3.03
	British White & Black African	1	3.03
	Total	**26**	**75.75**
Eastern European	Romanian	1	3.03
	Total	**1**	**3.03**
Non-European	Chinese	1	3.03
	Egyptian	1	3.03
	Filipino	1	3.03
	Japanese	1	3.03
	Malaysian	1	3.03
	Nigerian	1	3.03
	Vietnamese	1	3.03
	Total	**7**	**21.21**

Figure 4

Ethnic Composition in Sample Compared to London



Note. London data derived from England and Wales Census 2021 (Office for National Statistics, 2022).

Primary language spoken at home was also recorded, to explore whether those who have English as their preferred language were advantaged over those who do not. In this sample, 39% reported English as their first language, while 61% of the sample reported languages other than English as being their first language (see Table 5).

Table 5

Participant Primary Language Spoken at Home

Primary Language	Frequency	Percentage
English	13	39.39
Polish	7	21.21
Italian	2	6.06
Japanese	1	3.03
Arabic	1	3.03
Chinese	1	3.03
Greek	1	3.03
Filipino	1	3.03
Romanian	1	3.03
Slovakian	1	3.03
Spanish	1	3.03
Swahili	1	3.03
Vietnamese	1	3.03
Yoruba	1	3.03

Furthermore, as research has shown a link between parental SES and executive functioning skills, father occupation was recorded, and these were classified based on the National Readership Survey Social Grade system (National Readership Survey, 2006; see Table 6). Six children did not know their fathers' occupation and so this is recorded as "missing data" in Table 6. This data shows that the sample were diverse in terms of family SES; the most common social class background was Skilled Working followed by Lower-Middle and Middle-Middle Class and few children from other categories.

Table 6

Participant Fathers Occupation

Grade	Social Class	Occupation types	Frequency	Percent
A	Upper-Middle	Higher managerial	1	3.03
B	Middle-Middle	Intermediate managerial	7	21.21
C1	Lower-Middle	Junior managerial	7	21.21
C2	Skilled Working	Skilled manual workers	11	33.33
D	Working	Semi-skilled / unskilled manual	1	3.03
E	Non-working	State pensioners, unemployed	0	0.00
		Missing data	6	18.18

Note. Social class and grade classified based on the National Readership Survey (NRS) Social Grade system (National Readership Survey, 2006). Missing data was due to the child not knowing their fathers job role.

2.12.1. <u>Cognitive Data</u>

WISC-IV Similarities and Matrix Reasoning: participants were administered the Similarities and Matrix Reasoning subtests from the WISC-IV (Wechsler, 2003) to (a) identify any participants who may have visual or verbal difficulties, and (b) as a measure of concurrent validity. Table 7 shows participants age-scaled scores which standardises their raw score based on a normative sample of children of the same age in the UK (mean = 10 and standard deviation = 3). These scores suggest that this sample's verbal skills were better than the UK average which should be taken into consideration when interpreting the results of this study.

Visual skills, however, were closer to the UK average, though for both subtests there was a range in abilities. In this sample, Similarities SS ranged from eight (average) to 18 (very superior) indicating that no participants had verbal difficulties. For Matrix Reasoning, however, one child achieved a SS of only four, which can be classed within the "borderline" range. This score may be suggestive of visual processing difficulties or performance may have been impacted by other factors such as fatigue or attention. All other participants achieved SS between seven (low average) and 16 (very superior) in this subtest.

Mann-Whitney exact tests using the resampling procedures available in SPSS (IBM Corp, 2021) were used to check for group differences in WISC-IV Similarities and Matrix Reasoning scores based on sex and having English as a primary language (also shown in Table 7). It was found that Similarities scores did not differ significantly between those who have English as their first language compared with those who do not. Similarities scores also did not differ between males and females.

Table 7

WISC-IV and CHEXI age-scaled scores, and group differences based on Gender and English as a Primary Language.

Subtest	Mean SS	SD	English Language *U*	*z*	*p*	Sex *U*	*z*	*p*
WISC-IV Similarities	12.76	2.51	116.00	-.52	.62	120.00	-.58	.58
WISC-IV Matrix Reasoning	10.58	2.24	123.00	-.26	.81	105.00	-1.13	.28
CHEXI Inhibition	11.97	2.90	123.50	-.24	.82	95.50	-1.47	.14
CHEXI Working Memory	10.91	2.83	127.00	-.11	.92	107.00	-1.06	.30

Note. SD = Standard Deviation; *U* = Mann Whitney Test; *z* = *z*-score, *p* = exact significance.

Similarly, group differences were also not found for Matrix Reasoning scores between those who have English as their first language and those who do not, nor between males and females.

2.12.1.1. *Teacher-rated CHEXI:* Teachers were additionally asked to complete the CHEXI to explore any CF difficulties within the sample, and as a test of predictive validity. Scores were categorised into two domains: working memory and inhibition (Catale et al., 2015) and are also summarised in Table 7. Working memory was close to the Catale (2015) sample mean, however, inhibition SS was slightly above.

Mann-Whitney tests were used to check for group differences in CHEXI scores based on sex and English as a primary language (also shown in Table 7). CHEXI Inhibition scores did not differ between those who have English as their first language compared with those who do not. Scores were comparable, overall, for males and females. CHEXI Working memory scores also did not differ between those who have English as their first language and those who do not, and there were no group differences between males and females.

3. RESULTS

The following section will describe the exploratory analysis of the scoring procedure developed, to consider whether the measures are *feasible* methods of scoring performance in the Alien Game. It will then address the validity of these measures using correlational analysis. It will evaluate the Alien Game stimuli by exploring whether the individual attributes of the developed alien illustrations were identifiable; and consider the *acceptability* of the alien game by exploring the Likert-scale feedback from participants. Participant data was analysed using SPSS (Statistical Package for the Social Sciences) Version 28.0 (IBM Corp, 2021).

3.1. Exploratory Data Analysis

Exploratory data analysis was undertaken: histograms were generated as a visual check, in addition to descriptive statistics including skewness (skewness value of >1 identified as a skewed distribution), kurtosis (kurtosis values of >3 identified as a distribution with abnormal kurtosis), and Shapiro-Wilk's test ($p < 0.05$ identified as a non-normal distribution). Boxplots were also generated to check the data for outliers and/or data entry errors.

3.2. Alien Game Evaluation

3.2.1. Acceptability

Likert scale ratings ranged from one (awful), to five (fantastic), and results are given in Table 8. The most common game rating was four ("really good"), accounting for 64% of the sample, while the second most common rating was a five ("fantastic") accounting for 30% of the sample. Two participants (6%) scored the game a three (okay'); no participants rated the game a one ("awful") or a two ("not very good").

Table 8

Likert scale enjoyability rating for the Alien Game

Likert Scale rating	Frequency	Percentage
1	0	0.00%
2	0	0.00%
3	2	6.06%
4	21	63.64%
5	10	30.30%

Note. Likert scale rating: 1 = awful, 2 = not very good, 3 = okay, 4 = very good, 5 = fantastic.

3.2.2. Evaluation of Attributes

Analysis of the questions asked was split into three categories: those which are hypothesis-testing (HT, e.g., guessing which alien is the target), constraint seeking (CS, e.g., a conceptual question about a characteristic), and wrong format questions (those that cannot be answered with “yes” or “no”). Only 22 out of the 251 total questions asked were wrong format, which supports Pavitt’s (2017) findings that participants understood the task rules and approached the game in a strategic way.

Analysis of the constraint-seeking questions, summarised in Table 9, reveals that all 13 attributes were identified and asked about in every trial, suggesting they were all salient and distinguishable. Colour was asked about most, especially in the first two questions revealing that this was the most common strategy adopted to reduce the number of potential targets, followed by questions about eyes and legs. Most questions were consistent in format and directly referred to the attribute in the question, either asking if the attribute was present or whether a certain number of the attributes were present (e.g., does the alien have wings?, or “does the alien have two eyes”?). The only attribute which had less consistent questions was the antennae. While many participants did directly call them antennae, several participants were not familiar with the name and referred to these as “tubes,” “things,” or “plants,” while some pointed to them. This suggests that participants were able to distinguish and identify the antennae as an attribute though could not name them. In the future, it would be useful to specifically state in instructions that participants can point to characteristics if they unsure of the name.

Table 9

Frequency and percentage of questions asked about each alien attribute.

Attribute	Frequency of Questions	Percentage of Questions
Colour	200	26.18
Eyes	107	14.01
Legs or feet	107	14.01
Wings	72	9.42
Eyebrows	45	5.89
Tail	40	5.24
Antennae	37	4.84
Nose	37	4.84
Horns	36	4.71
Shape	31	4.06
Teeth or fangs	31	4.06
Ears	11	1.44
Arms or hands	10	1.31

Additionally, some composite questions were asked about more than one attribute (e.g., does the alien have legs and wings?), and some asked about the position of the target on the board (e.g., is it in the first row?). In this study, these questions were permitted; however, the instructions could be adapted to explain whether these types of questions are permitted or not.

3.3. Alien Game Performance

Descriptive statistics for the measures which were trialled to evaluate performance in the Alien Game are summarised in Table 10, and SPSS tables and histograms are presented in Appendix S. Measures with non-normal skewness, kurtosis, SD, or Shapiro-Wilk values are in bold font.

Table 10

Descriptive Statistics for Novel Alien Game Measures

Measure	Mean	SD	Min.	Max.	Skewness (SE = .41)	Kurtosis (SE = .80)	Shapiro-Wilk test	
							W	*Sig*
TQ's	28.21	8.34	14.00	49.00	.69	-.05	.95	.10
WA	12.06	4.81	3.00	19.00	-.41	-.55	.94	.07
IA	32.12	5.75	11.00	48.00	**-1.05**	**6.34**	.78	**<.01**
Ab	65.15	7.07	45.00	76.00	-.54	.63	.96	.23
IU	11.91	8.42	1.00	31.00	.69	-.50	.93	**.03**
WIU	7.48	4.55	.00	15.00	-.03	-1.02	.96	.19
AGS	19.06	9.64	.00	34.00	-.34	-.75	.95	.16

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

3.3.1. Total Questions (TQ)

This measure was calculated by adding together questions asked across the four trials (A-D). If the best possible strategy is adopted, the target alien should be identified in the fewest possible number of questions; therefore, for TQ measure, better performance is indicated by a smaller value.

Boxplots (Figure S1) reveal a slightly longer tail of higher scores compared with lower scores, however no values were identified as outliers. The histogram (Figure S2) reveals a good range of scores, with a bell-shape curve which does not significantly depart from normal.

TQ was therefore identified as a feasible measure of performance in the Alien Game, as it is easy to calculate, produces a good range of scores with no floor or ceiling, and scores have a normal distribution.

3.3.2. Weighted Achievement Score (WA)

The WA score was calculated by grading TQ on a scalar (see Table 1), where a higher WA score indicates better performance. The boxplot (Figure S3) reveals no outliers, with a slightly longer tail for lower scores compared with high scores. The histogram (Figure S4) and Shapiro-Wilk test reveals that the distribution does not depart from normal. This measure has a good range of scores with no ceiling or floor.

WA was therefore identified as a feasible measure of performance in the Alien Game as it produces a good range of scores and a normal distribution. Note, however, that in the future, the point system may require adjusting based on a larger sample of normative data.

3.3.3. Initial Abstraction Score (IA)

The IA was calculated by summing the number of targets that could be removed by the first question across the four trials; higher scores indicate better performance.

The boxplot (Figure S5) reveals a small distribution of data, with a few scores identified as outliers. The histogram (Figure S6) suggests leptokurtic distribution centred around a score of 30, with few scores on the tails of the distribution. Shapiro-Wilk test confirmed that the distribution departs from normal.

Further exploration of the data revealed that the possible number of targets which could be removed by the first question is very limited, as most questions will either remove a third of options, half of options, or a single target. Most participants asked a question that removed 33% of targets in their first question (that is, the colour of the alien). As there is little variation in the scores, IA therefore appears to not be a useful measure of performance in the Alien Game.

3.3.4. Abstraction Score (Ab)

Ab was calculated by summing up the minimum number of targets removed by the first three questions asked across the four trials, where a higher score means better performance. This created a good range of scores (Figure S7) in a bell curve (Figure S8) which appeared slightly negatively skewed due to one low score. Shapiro-Wilk and skewness and kurtosis tests reveals that the distribution does not depart from normal distribution.

This measure, therefore, was identified as a feasible measure of performance in the Alien Game, yielding a good range of scores within a normal distribution.

3.3.5. Ineffective or Unallowed Questions (IU) Score

A question was considered as ineffective or unallowed if it failed to remove any items, was a repeated question, a wrong guess (i.e., a HT question) or a question that could not be responded to with a “yes” or “no” answer. Total IU questions therefore encapsulates the number of these questions over the four trials, where a lower figure indicates better performance.

The measure yielded a good range of scores, a long tail for higher values compared to lower values, but no outliers (Figure S9). Although the histogram (Figure S10) appears non-normal, Shapiro-Wilk test suggests that the distribution does depart from normal. Skewness and kurtosis were found to be in the normal range. Overall, this measure was found to be a feasible measure of performance in the Alien Game as a good range of scores, while the distribution of these scores was non-normal in this sample.

3.3.6. Weighted Ineffective or Unallowed Questions (WIU) Score

Weighted ineffective questions score provides points based on the number of ineffective questions, and a higher score means better performance. The boxplot (Figure S11) reveals no outliers, and a good range of scores. The histogram (Figure S12) appears somewhat flat, but Shapiro-Wilk test, kurtosis and skewness suggests that the distribution does not depart from normal. This was identified as a feasible measure for The Alien Game as a good range of scores were produced which fit within a normal distribution.

3.3.7. Alien Game (AG) Score

The overall AG Score is calculated by summing the WIU and the WA score and so provides a measure which incorporates the number of questions asked with the quality of questions. Higher scores in this measure indicates better performance.

A good range of scores were produced. The boxplot (Figure S13) reveals a longer tail for lower scores compared with higher scores, though the histogram (Figure S14) shows a bell-shaped curve. Shapiro-Wilk test, kurtosis and skewness reveals that the distribution does not depart from normal. AGS therefore appears to be a feasible measure of global performance.

3.3.8. Summary

Six of the seven measures we developed were identified as feasible methods to evaluate performance in the Alien Game. These included total questions (TQ), weighted achievement score (WA), abstraction score (Ab), ineffective or unallowed questions (IU), weighted ineffective or unallowed questions (WIU), and Alien Game Score (AGS). Each reveal a good range of scores with no ceiling or floor effects, while remaining user friendly to calculate. Initial abstraction score was not a useful measure.

3.4. Validity

As some of the variables were shown to be non-normally distributed, Spearman's rank correlations were used as non-parametric tests for analysis. Firstly, we looked at the relationships within the Alien Game measures to explore whether these measure separable skills. Then we addressed the relationship between the WISC, CHEXI, and the Alien Game measures to explore concurrent ad predictive validity. Mann-Whitney tests were used to explore the influence of age, sex, and English language on test performance.

3.4.1. Correlation Between Alien Game Measures

The correlations between Alien Game test performance measures are summarised in Table 11. It may be seen that measures TQ, IU, WA, WIU and AGS were all strongly associated with each other. In contrast, Ab exhibited a moderate association with TQ, but no relationship to any of the other measures.

Table 11

Spearman's Rank Correlations within the Alien Game Measures

Measure	TQ		Ab		IU		WA		WIU	
	r_s	Sig	r_s	Sig	r_s	Sig	r_s	Sig	r_s	Sig
Ab	**.36**	**.04**								
IU	**.94**	**<.01**	.23	.19						
WA	**-.90**	**<.01**	-.16	.36	**-.89**	**<.01**				
WIU	**-.90**	**<.01**	-.23	.20	**-.96**	**<.01**	**.84**	**<.01**		
AGS	**-.92**	**<.01**	-.21	.24	**-.96**	**<.01**	**.94**	**<.01**	**.97**	**<.01**

Note. r_s = Spearman's rank correlation coefficient*;* p = exact significance level. Moderate associations (r_s >.3) and strong associations (r_s > .5) are shown in bold.

3.4.2. Correlation Between Concept Formation Measures

The correlations between the WISC-IV and CHEXI subtests are summarised in Table 12. A positive correlation was found between Matrix Reasoning and Similarities and between CHEXI Working Memory and Inhibition. This was expected as these measures are part of the same instrument. There was no correlation between Similarities and CHEXI Inhibition or Working Memory, nor between Matrix Reasoning and the CHEXI subscales.

Table 12

Spearman's Rank Correlations between Concept Formation Measures

Measure	WISC Similarities		WISC Matrix Reasoning		CHEXI Inhibition	
	r_s	Sig	r_s	Sig	r_s	Sig
WISC Matrix Reasoning	**.39**	**.02**				
CHEXI Inhibition	-.12	.50	-.15	.41		
CHEXI Working Memory	.07	.70	.19	.29	**.67**	**.00**

Note. Moderate associations (r_s >.3) and strong associations (r_s > .5) are shown in bold.

3.4.3. Concurrent validity

To assess concurrent validity (i.e., whether scores in the Alien Game are associated with established measures of CF), we conducted correlation analysis comparing the Alien Game measures to Matrix Reasoning and Similarities, shown in Table 13. A negative correlation was found between TQ and Matrix Reasoning. No associations were found between performance measures and Similarities or CHEXI subscales.

Table 13

Spearman's Rank Correlations between Alien Game Measures, WISC-IV and CHEXI subscales.

Measure	WIS Similarities		WISC Matrix Reasoning		CHEXI Inhibition		CHEXI Working Memory	
	r_s	Sig	r_s	Sig	r_s	Sig	r_s	Sig
TQ	.02	.91	**-.35**	**.05**	.12	.51	-.07	.68
Ab	.09	.63	-.16	.38	.17	.35	.15	.41
IU	-.03	.85	**-.34**	**.06**	.01	.97	-.17	.35
WA	-.05	.77	.25	.12	-.06	.75	.09	.62
WIU	.09	.62	**.33**	**.06**	-.07	.69	.16	.36
AGS	.02	.90	.29	.10	-.05	.76	.16	.38

Note. Moderate associations (r_s >.3) are shown in bold.

3.4.4. Predictive validity

To assess predictive validity (i.e., whether scores on the Alien Game correlate with real-world executive functioning) we compared Alien Game performance to scores obtained in the CHEXI. As shown in Table 13, no associations were found between Alien Game performance measures and CHEXI subscales.

3.4.5. Influence of demographic data on Alien Game performance

Mann-Whitney tests were used to explore influence of age, sex, or English as a primary language on Alien Game performance measures. As summarised in Table 14, no group differences were found for any of the Alien Game performance measures.

Table 14

Alien Game Performances Grouped by Age, Sex and English as Primary Language

			Mean	SD	U	*z*	Sig
TQ	English as Primary Language	Yes	27.08	8.14	110.00	-.74	.47
		No	28.95	8.59			
	Age	8-9	27.50	8.79	120.50	-.56	.59
		10-11	28.88	8.10			
	Sex	Male	29.81	8.92	111.00	-.90	.38
		Female	26.71	7.72			
Ab	English as Primary Language	Yes	64.38	8.20	129.00	-.04	.98
		No	65.65	6.40			
	Age	8-9	65.00	5.54	124.00	-.43	.68
		10-11	65.29	8.43			
	Sex	Male	64.44	7.63	122.00	-.51	.62
		Female	65.82	6.65			
IU	English as Primary Language	Yes	10.46	8.18	106.50	-.87	.40
		No	12.85	8.64			
	Age	8-9	11.56	9.04	123.00	-.47	.65
		10-11	12.24	8.05			
	Sex	Male	13.06	8.84	114.50	-.78	.45
		Female	10.82	8.12			
WA	English as Primary Language	Yes	12.77	5.54	110.00	-.74	.47
		No	11.60	4.37			
	Age	8-9	12.00	5.23	135.50	-.02	.99
		10-11	12.12	4.55			
	Sex	Male	11.44	4.95	116.00	-.72	.48
		Female	12.65	4.76			
WIU	English as Primary Language	Yes	8.00	5.12	114.00	-.59	.57
		No	7.15	4.25			
	Age	8-9	7.88	7.12	122.00	-.51	.62
		10-11	5.03	4.17			
	Sex	Male	6.38	4.56	96.50	-1.43	.16
		Female	8.53	4.42			
AGS	English as Primary Language	Yes	20.77	10.41	110.50	-.72	.48
		No	18.75	8.20			
	Age	8-9	17.81	9.25	108.00	-1.01	.32
		10-11	21.18	8.78			
	Sex	Male	19.88	9.98	130.50	-.20	.85
		Female	19.24	8.35			

Note: *M* = mean; SD = Standard Deviation; U = Mann Whitney Test; z = z-score; *p* = exact significance.

3.4.6. Validity summary

Correlations were found between performance in all measures, apart from Ab which only significantly correlated with TQ's. This suggests that Ab may be measuring a different aspect of the test to that of the other measures. It was also found that performance in the Alien Game is not related to expressive executive function skills as measured by CHEXI Inhibition or CHEXI Working Memory. Additionally, performance measures did not correlate significantly with performance on WISC-IV Similarities, suggesting performance in the Alien Game is not related to verbal skills. Some Alien Game performance measures do, however, significantly correlate (or are approaching significance) with Matrix Reasoning performance. This suggests that performance in the game may be related to nonverbal receptive abstraction skills, and this also provides evidence of concurrent validity. Lastly, it was found that gender, age and having English as a primary language do not influence performance on the Alien Game in this sample.

3.5. Reliability

Reliability was explored by looking at change in scores from trial A to trial D for TQ, Ab and IU. Wilcoxon Signed Ranks Tests were used to compare performance in these measures, and these are summarised in Table 15

A significant difference in performance was found from trial A to trial D for Ab only, suggesting performance was better in trial D compared to trial A in this measure. Exploration of the descriptive statistics, however, shows performance was changeable across the four trials. TQ and IU did not show a significant change in score from trial A to trial D. Overall, therefore, reliable improvements in performance from trials A-D were not found.

Table 15

Descriptive Statistics and Wilcoxon Signed Ranks Test comparing Performance Measures across Trials

	Mean	SD	Min.	Max.	*z*	*p*
Total Questions Trial A	6.64	3.00	3.00	17.00		
Total Questions Trial B	7.33	3.24	3.00	14.00		
Total Questions Trial C	6.33	2.84	3.00	13.00		
Total Questions Trial D	7.91	*3.47*	*4.00*	*18.00*		
Total Questions Trial A – Trial D					-1.80	.07
Abstraction Score Trial A	15.91	*4.21*	*3.00*	*21.00*		
Abstraction Score Trial B	16.24	*3.43*	*8.00*	*24.00*		
Abstraction Score Trial C	15.18	*3.48*	*9.00*	*20.00*		
Abstraction Score Trial D	17.80	*2.24*	*12.00*	*20.00*		
Abstraction Score Trial A – Trial D					***-2.50***	***.01***
Ineffective Questions Trial A	2.85	*3.11*	*.00*	*12.00*		
Ineffective Questions Trial B	4.03	*4.30*	*.00*	*19.00*		
Ineffective Questions Trial C	2.79	*2.91*	*.00*	*10.00*		
Ineffective Questions Trial D	3.73	4.79	.00	21.00		
Ineffective Questions Trial A – Trial D					*-.36*	*.72*

Note. Z = z score; *p* = exact significance; significant z-scores at 95% confidence level are shown in bold.

4. DISCUSSION

This section will revisit the research aims and questions and then discuss these in relation to the research findings. The findings of this study will then be situated within the literature presented in the introduction. A critical review of the study's strengths and limitations will then be discussed. The clinical implication of the findings will then be explored, followed by suggestions for future research.

4.1. Summary of Research Questions and Aims

The aim of this study was to develop the Alien Game which was previously piloted by Pavitt (2017) where it was found to have potential as a test of concept formation (CF) in children. The current study then aimed to run a second pilot study with children aged 8-11 years and to then develop a scoring procedure. The research questions were as follows:

1. Can a culturally fair test of concept formation be derived from the Alien Game, for children aged 8-11 years?
2. Will the test format be acceptable and engaging for children?
3. To what extent will Alien Game scores correlate with established measures of concept formation, such as WISC-IV Matrix Reasoning and Similarities?
4. To what extent will the Alien Game scores correlate with objective reports of executive functions in children, such as in the CHEXI?

4.2. Summary of Research Findings

4.2.1. <u>Research Question One</u>

Following the first pilot of the original Alien Game (Pavitt, 2017), improvements to the materials, stimuli and scoring protocol were suggested in order to develop the Alien Game as a fair test of CF for children aged 8-11 years. This question will be broken down into three parts to explore how the current study addressed these suggestions by improving and refining the materials and developing a scoring protocol, and lastly how the Alien Game addresses CF in a culturally fair way.

4.2.1.1. Improving and refining materials: Firstly, novel stimuli were used to eliminate possible pre-exposure effects, and the alien illustrations were developed so that the individual attributes were clearer and easier to distinguish. Additionally, the number of aliens assigned to each attribute category was evened out, so each category contained the same number of aliens. For example, 12 aliens had wings and 12 did not; the aim of this was to reduce the element of chance and to make the Alien Game a more stringent test of CF. Lastly, the plastic format of the game was removed to reduce cost and to make the game more challenging. The aim of this was also to improve ecological validity to reflect real-world demand.

To assess these changes, the questions asked by participants were explored. It was found that participants asked questions about all 13 characteristics of the aliens, suggesting they were all identifiable and distinguishable. Furthermore, it was found that most participants used the same approach of asking constraint-seeking questions using attribute categories to eliminate groups of aliens with shared features. Very few questions were in the wrong format (for example, questions which could not be answered with ‘yes’ or ‘no’), which provides support for the use of this game as an assessment tool. Additionally, nearly all participants were able to identify the target alien using the printed grid, where targets could not be physically removed, providing support for the use of this format with children in this age group. Overall, these findings suggest that the developments made to the Alien Game are accessible to children aged 8-11 years.

4.2.1.2. *Developing a scoring protocol:* the number and types of questions participants asked were explored to derive potential approaches to measure performance in the Alien Game. The resulting data was analysed using exploratory analysis and descriptive statistics, to determine whether there was sufficient typicality and variability in performances, and whether this was captured by the measures developed. The distribution of the data was explored visually using histograms and boxplots, and statistically using Shapiro-wilk, and skewness and kurtosis measures. Spearman’s rank correlations were used to investigate the relationships between these measures. The performance measures identified include Total Questions (TQ), Weighted Achievement Score (WA), Abstraction Score (Ab), Ineffective or Unallowed Questions (IU), Weighted Ineffective or Unallowed

Questions (WIU) and Alien Game Score (AGS). Initial Abstraction (IA) Score on the other hand was not found to be a good measure of performance.

The results indicate that these six measures yielded a good range of scores, with no ceiling or floor, suggesting they successfully measure variation in performance. The Total Questions measure, however, was found to correlate with Ineffective or Unallowed Questions (and WIU's), suggesting that they are measuring similar aspects of performance: children who needed more questions to identify the correct target were also likely to ask more ineffective or unallowed questions. It may therefore not be necessary to include both as indices of performance going forwards, though the Ineffective and Unallowed questions measure may provide qualitative insight into area of difficulty.

These measures (TQ, IU and WIU) were also the three most closely correlated with Matrix Reasoning scores. This suggests that total questions asked provides a good measure of performance and may reflect nonverbal abstraction skills. Notably, Abstraction Score correlated with TQ, but did not correlate with Matrix Reasoning or other performance measures. This suggests that Abstraction Score, while hypothesised to measure abstraction, may be measuring a separate aspect of performance.

Weighted Achievement Score and Alien Game Score were also highly correlated with TQ, IU, and WIU, which was expected as these two measures use the same data to capture overall performance. Alien Game Score was developed to capture overall performance and was calculated in this study using Weighted Achievement Score and Weighted Ineffective and Unallowed questions. As these two measures appear to be capturing similar skills, Alien Game Score could be developed to capture total questions and abstraction score instead which appear to be measuring different aspects of performance.

Overall, the findings suggest that the measures developed successfully capture variation in performance, providing support for the Alien Game as an assessment tool. The scoring protocol is further discussed in section 4.3.

4.2.1.3. *The Alien Game as a culturally fair test:* Tests of CF, such as those used by Piaget (1952) to investigate childhood cognitive development, are now thought to advantage those who have grown up in the West and who speak English as their first language. This is because these tests include Western concepts and require proficiency in English for administration and response. It is important, therefore, to have available a nonverbal assessment of conceptual skills, which meets the needs of all children, regardless of their background, culture and languages spoken. We aimed to achieve this by recruiting a diverse sample, and by reducing the language requirements of the Alien Game as much as possible to produce a visual test of abstraction. Adaptions included removing the alien's names from the stimuli so only perceptual features were used as categories, as well as the use of novel images to avoid familiarity with the propriety game "Guess Who?" advantaging some participants. The original Guess Who game stimuli are also culturally laden and uses faces and clothing as categories.

Descriptive statistics suggest that this sample was representative of the ethnic composition of London, and was diverse in ethnicity, first language, sex, and SES. Most children came from Western-European backgrounds, but many were non-European, and two-thirds of the sample reported a primary language other than English. The sample represented a fairly even spread of children from working class, middle-class and upper-class SES groups, as well as an even number of males and females. The sample is reflective of diversity within London-based schools, which supports the generalisability of the data.

There was no relationship between performance in the Alien Game and Similarities, suggesting that performance in the Alien Game is not related to verbal skills. There were also no differences between male and female participants, nor between those who speak English as a first language and those who do not, suggesting that performance is not influenced by language or sex in this sample. Group differences based on English as a first language, however, were also not found for Similarities performance. Average performance in the Similarities subtest was above the UK average in this sample which potentially masked differences in Alien Game performance due to first language. Nevertheless, there was a range of verbal ability in the sample, from average to very superior, and so a correlation with Alien Game performance would still be expected if language abilities influenced performance. Future studies could explore this finding further and aim to recruit a sample of children with a greater range of verbal abilities.

Performance in the Matrix Reasoning subtest was closer to the UK average, suggesting the sample was more reflective of the UK population in nonverbal abilities. A negative correlation was found between this measure and total questions asked, suggesting that performance taps into nonverbal abstraction, though not verbal abilities, supporting the use of this game as a culturally fair assessment of CF.

4.2.2. Research Question Two

The second aim of the study was to create a test which will be engaging for children aged 8-11. As noted earlier, tests designed for adults are often not engaging for children, potentially leading to fatigue, boredom, and reduced attention; while very formal assessment formats may lead to anxiety (Berg et al., 2020; Kim et al., 2020), which may negatively impact performance. Research has also shown that tasks which are enjoyable and interesting lead to better performance (Schukajlow & Krug, 2014). It is therefore important to create tests specifically for children that are engaging and targeted to their developmental stage.

The Alien Game included child-friendly illustrations, and a game-like format to appeal to children and to maintain engagement. The cartoon stimuli were made child-friendly by keeping them simple and using colourful illustrations. The information sheet, consent form, and debrief form were also designed to be colourful and use simple English to be accessible and engaging to children. Child-friendly language was used throughout, and attention was paid to how the tasks were introduced and explained, to minimise test anxiety.

Engagement with the test was explored by careful inspection of the performance measures and questions asked. We compared performance by age group to explore whether adaptations to the test would need to be made for children in different age groups. Feedback was collected in the form of a Likert-scale to confirm that children enjoyed the game, and observational data was also used.

It was found that children engaged well, and feedback confirmed that children enjoyed the game. The game was found to meet the developmental age of children as performance measures yielded a good range of scores with no ceiling or floor, children were able to identify all 13 attributes of the aliens, and children approached the task in a consistent manner. Furthermore, no differences in performance were found between children aged 8-9 and 10-11 suggesting the task is suitable for children across these ages.

Overall, these findings suggest the game-like elements of the Alien Game were engaging to children, and the task was not too easy or difficult for children of this age. Child-friendly assessments of CF further discussed in section 4.5.

4.2.3. Research Question Three

The third research question was concerned with how scores on the Alien Game related to established measures of CF. Negative correlations were found between WISC-IV Matrix Reasoning and both Total Questions and Ineffective or Unallowed Questions, while a positive correlation was found between Matrix Reasoning and Weighted Ineffective or Unallowed Questions, providing evidence of concurrent validity. On the other hand, Alien Game performance scores were not related to Similarities. These findings suggest that performance in the Alien Game is not underpinned by verbal skills, and performance is more closely related to nonverbal abstraction skills.

4.2.4. Research Question Four

Question four was concerned with predictive validity of the Alien Game. This was measured by comparing performance in the Alien Game to CHEXI measures of inhibition and working memory, however no correlations were found. While predictive validity was therefore not established, this finding may be related to differences in administration of the Alien Game and the CHEXI. It may also suggest that abstraction is not underpinned by inhibition and working memory. This finding is discussed further in section 4.4.

4.2.5. Summary of Findings

In summary, the scoring methods developed were found to be feasible approaches to measure performance in the Alien Game for children ages 8-11 years. The scoring methods are user-friendly and easy to calculate and were found to correlate with an established nonverbal measure of abstraction providing evidence of concurrent validity. On the other hand, performance was not found to be related to verbal skills, which along with the use of novel images, supports the use of the Alien Game as a more culturally fair assessment of CF compared with those which are currently available for children. This was further supported by the finding that performance in the Alien Game as not impacted by having English as an additional language. Predictive validity, however, was not established; so further research is required to explore this. Importantly, participant feedback was positive, and observations of the children playing the game showed they engaged well and appeared to enjoy taking part.

4.3. Exploration of the Scoring Procedure

4.3.1. Total Questions

This measure is the sum of the number of questions required to identify the target and was designed to capture performance broadly. This measure was influenced by the “Total Question Asked” score used in the DKEFS 20 Questions test (Delis et al., 2001) as a measure of global achievement. The authors recommend developing a Weighted Achievement Score from the total questions asked, as a lucky guess may result in requiring fewer questions to identify the target, which may not be representative of the participants abilities.

We found total questions to be a good measure of variation in performance, and it was also the measure most closely related to established measures of abstraction, suggesting it is capturing concept formation abilities. Summing the total number of questions required to reach the target is also easy to calculate for busy clinicians.

Alderson-Day & McGonigle-Chalmers (2011) Robot Task used a similar format to the Alien Game to measure CF. They did not use total questions asked as a global measure of performance and instead calculated the percentage of trials “completed” (target identified within 10 questions), along with additional measures to explore strategy more closely. This scoring method, however, is more difficult to calculate.

A disadvantage of using total questions as a performance measure is that many different skills are likely required to identify the target in minimal questions, including CF, problem-solving, working memory, attention, and concentration. Therefore, a high total questions score does not provide insight into what the participant struggled with, and so further measures can provide insight into strategy and specific difficulties. Additionally, as mentioned previously, a low total questions score may be the result of a lucky guess.

4.3.2. Weighted Achievement Score

The limitation of the total questions score being artificially reduced by a lucky guess was addressed by developing a Weighted Achievement Score. A variation of this was also used in the DKEFS 20-Questions task so that those who fortuitously guess the correct target using only a few concrete questions would not be overlooked. This score provides points on a u-curve, where those who arrive at the correct answer in too few or too many questions are awarded fewer points than those who ask the optimal number of questions. In the DKEFS 20-Questions task, the optimal number of questions required to reach the target was identified from the norm data, and so further development of the Weighted Achievement Score in the Alien Game may be required following collection of additional data.

4.3.3. Ineffective or Unallowed questions

This measure was developed following exploration of the types of questions participants asked. The question types classed as ineffective or unallowed include questions not in a yes/no format, repeated questions, questions which fail to eliminate any targets, and wrong guesses. This measure was found to correlate with the number of questions asked as those who ask more ineffective questions, or questions which break the rules of the game, are likely to require more questions overall to identify the target. Both measures therefore may not be necessary, though this measure may still be useful for qualitative purposes, to explore possible areas of difficulty. For example, if questions are repeated frequently, this may be due to working memory difficulty or perseveration, whereas if hypothesis-testing questions are most often used, this may suggest difficulties in abstraction or finding a good strategy. Alderson-Day & McGonigle-Chalmers (2011) explored question type in their similar test of CF and found children with ASD asked more hypothesis-testing questions, compared with typically developing children. They also found that this group difference increased when the ability to physically eliminate non-targets was removed. This suggests that exploring quality of questions may be useful for identifying the area of difficulty, especially when testing the Alien Game on clinical populations. The DKEFS 20 Questions task also incorporates a similar Error Score as an optional process in addition to the primary measures, so the Alien Game could also consider the ineffective or unallowed measure as an optional extra measure.

4.3.4. Abstraction Score

This measure was designed as a measure of abstraction and CF. This measure was influenced by the Initial Abstraction measure used in the D-KEFS 20 Questions test (Delis et al., 2001), calculated from the number of targets removed by the first question only. The authors suggest that this score represents problem-solving efficiency and abstract thinking. In this study, we also developed an Initial Abstraction measure based on the number of targets removed by the first question only, however, the range of possible targets removed by the first question is limited in the Alien Game, and so this was not found to be a useful measure.

The Abstraction Score was instead calculated by adding together the total number of targets removed by the first three questions. In the Robot Task (Alderson-Day & McGonigle-Chalmers, 2011), a similar score was calculated across all 10 questions, using proportion of targets removed rather than the absolute number. In contrast, we found that summing the actual number of targets removed from the first 3 questions only, was sufficient and more user friendly.

In the Alien Game, therefore, Abstraction Score was hypothesised to measure the same skills at the DKEFS Initial Abstraction measure: the ability to identify the attribute category which can differentiate the largest proportion of remaining targets. We found that the Abstraction Score correlated with the number of questions asked, though did not correlate with any other performance measures, suggesting it may be measuring a different aspect of performance to the other measures. The Abstraction

Score also did not correlate with any pre-existing tests of abstraction, suggesting it may not be measuring abstraction skills as assumed. Rather, the measures Total Questions and Ineffective and Unallowed Questions were found to correlate with Matrix Reasoning scores, suggesting they may be most accurately measuring nonverbal abstraction. Future studies could explore this further, to see if these findings are replicable in different samples, and tests of other executive functions could be used to further explore what skills these two measures are capturing.

4.3.5. Alien Game Score and Weighted Ineffective or Unallowed questions

Lastly, the Alien Game Score was developed as an overall measure of performance. We created this measure by combining the total questions required to reach the target and the number of ineffective and unallowed questions asked, as these two measures were hypothesised to capture different aspects of performance. The transformed weighted score of these two measures were used, rather than the raw score, for ease as they could be added together. Following the finding that these two performance measures are strongly related, whereas abstraction score appears to be capturing a distinct aspect of performance, future research could develop this measure to represent abstraction score and total questions instead. This would also remove the need for the Weighted Ineffective or Unallowed questions measure (WIU), further improving ease of scoring.

4.4. Concept Formation

This next section will discuss the research findings in relation to the literature on the assessment of CF discussed in the introduction.

4.4.1. Relationship With Other Executive Functions

In this study, we found that inhibition and working memory measured by the CHEXI was not related to verbal and nonverbal abstraction as measured by the WISC-IV. It was also found that performance in the Alien Game was not related to inhibition or working memory, even though working memory was hypothesised to support performance as this is a multi-trial test. In contrast with previous theories of executive functions (Teuber, 1972), these findings suggest that receptive executive function skills, such as abstraction, may not be underpinned by inhibition and working memory.

Some research suggests that individual executive functions are strongly correlated with one another, for example according to Best et al (2009), the unity-and-diversity view suggests that individual executive functions are distinct but closely-related skills which have a common underlying mechanism. Other studies, however, have found more modest correlations between disparate executive functions, and Wiebe et al. (2008), found evidence of working memory and inhibition as separate cognitive functions. Research has also suggested that the developmental trajectory may be different for each executive function: for example, Isquith et al. (2004) suggest that inhibition may develop before planning and working memory skills. Our findings may support this view of executive functions as disparate skills which may develop independently.

Another explanation for the lack of correlation between measures may be the differences in administration (for example, the Alien Game and WISC measures are structured task versus the CHEXI which is a behaviour rating). Behaviour ratings are assumed to be more ecologically valid than structured tests as they measure real-world application; however, they are also highly subjective. Context is also missing from such measures. Behaviours noticed by teachers in the classroom (for example, students becoming fidgety and inattentive) may be attributed to executive function difficulties rather than the child feeling confused, anxious, or overwhelmed due to difficulties learning; or feeling bored and disengaged due to not being challenged by the work. Furthermore, children learn how to regulate behaviour over the course of development, but how well they do so may be contextual and situational. This may explain why children perform well in structured tests, and yet encounter difficulties in the classroom (Bernstein & Waber, 2007). Future studies could explore predictive validity further, by using different measures to capture real-world executive functioning for example, CF skills measured by parents.

Additionally, the WISC measures differ in format to the Alien Game as they are single trial measures of CF whereas the Alien Game is a multi-trial test, which may contribute to the lack of correlation between performances in these tests. Multi-trial tasks require multiple cognitive demands in addition to CF skills, (Smidts et al., 2004), though offer more ecological validity. There are widely available multi-trial tests for children which could have been used instead to establish concurrent validity, though they would have all presented different challenges. For example, the DKEFS 20 Questions test is very similar to the Alien Game in format which may have resulted in practice effects. Alternatively, the DKEFS Word Context test is also a multi-trail test though relies heavily on verbal skills, as does the DKEFS Proverbs subtest. The Proverbs test is also not suitable for children below age 16 and includes some high level and taxonomic concepts (for example, living versus non-living). Concurrent validity is therefore difficult to establish as the pre-existing measures may not be equivalent.

4.4.2. Learning Conceptual Skills

Rich learning opportunities and stimulating environments have been found to be important factors in the development of conceptual skills (Ford et al., 2019). Children from disadvantaged backgrounds have therefore been found to have less developed conceptual skills compared to advantaged peers. This suggests, however, that with early identification of such difficulties, interventions, and targeted learning plans, can support children from disadvantaged backgrounds to develop this important skill (Wilson, 2004).

As discussed in the introduction, previous studies which have explored teaching of conceptual skills, through training (Tzuriel & Klein, 1985) and repetition of completing conceptual tasks (Pasnak et al., 2006), have found improvements in skills over time. In contrast, we did not find evidence of consistent improvement in performance across the four trials of the Alien Game, suggesting learning did not occur. Studies which have found learning to occur, did so following weeks or months of intervention. This may suggest that four trials of the Alien Game may not be enough for improvement in conceptual skills to be seen, though learning may have been seen following additional trials of the Alien Game. This provides further evidence that while conceptual skills can be taught, children may require a prolonged intervention for meaningful change to occur.

4.4.3. Assessment of Concept Formation

Conceptual skills develop over childhood, and according to Piaget (1952), the children in this study sit within the concrete operational stage of cognitive development which lasts between 7-and-11 years of age. During this stage of development, conceptual thinking abilities are thought to be limited to concrete categories rather than abstract thought (Freyberg, 1966). Piaget suggested that children younger than 7 years do not have capacity for abstract thinking. However, weaknesses in Piaget's approaches and methods mean that young children's true capacity for abstract thought is greater than has traditionally been allowed and different results have been found in recent attempts to replicate his experiments with child-friendly adaptations (Borke, 1975). Accordingly, there is need for child-friendly assessment methods which reflect the developmental age of children for accurate assessment of abilities. This study found support for the use of the child-friendly adaptations used in the Alien Game. Some of the research presented in the introduction found evidence that children as young as 3 years of age showed evidence of emerging conceptual skills, and so future research could explore whether the Alien Game could be used with younger children. Adaptations may include limiting verbal requirements, targeting the child's developmental stage and considering the format and materials of the test.

4.4.3.1. *Limiting verbal requirements:* Existing tests of CF presented in the literature review often used visual stimuli or relied on perceptual categories to reduce verbal demands. However, tasks such as the Tinker Toy Test (Roberts et al., 1995), included additional verbal demands, as participants were required to describe the item they created, and a relationship was consequently found between performance and verbal skills. Condy (2021) also relied on semantic categories, as did Alderson-Day & McGonigle-Chalmers (2011) in their standard condition, and therefore performance in these tasks is also likely related to verbal abilities. In contrast, in this study, while children were asked to respond verbally, the quality of their verbal answers was not germane.

Instead, the Alien Game used only perceptual categories, such as colour or number of attributes. The number of categories was reduced from the previous version of the Alien Game, to make the images clearer and easier to identify. Alt et al. (2013) found primary school age children may find abstracting category membership difficult as children aged 6-9 years used concrete categories to organise material, and were less forgiving of atypical examples. The stimuli used in the Alien Game, was found to be accessible and engaging to children aged 8-11 years.

Consequently, performance in the Alien Game was found to not be related to verbal skills. Nonverbal measures are more culturally fair and are more child friendly, and the Alien Game provides support for the feasibility of a CF assessment which does not require verbal responses.

4.4.3.2. *Considering the Format and Materials of the Test:* Task adaptions included in previous studies which aimed to improve engagement have consequently been too difficult or too easy for children and have therefore not been appropriate to the developmental stage of the child. For example, Roberts et al (1995) found the Tinker Toy task was not sensitive enough to distinguish performance in normally developing children, and a ceiling effect was also found in Alderson-Day & McGonigle-Chalmers (2011) Robots Task. On the other hand, use of technology proved too challenging for a clinical population in Condy et al.'s (2021) task. In contrast, the Alien Game used simple stimuli and was found to successfully measure variation in performance.

Some studies have attempted to improve enjoyability in their assessments by using stimuli that were engaging to children, such as computer or tablet-based assessments (for example, Alt et al., 2013; Condy et al., 2021), or the use of plastic toys or apparatus (Alderson-Day & McGonigle-Chalmers, 2011; Roberts et al., 1995; Smidts et al., 2004). However, in some studies this also impacted the accessibility and affordability of the assessments, and use of technology was also found to be too challenging, which negatively impacted engagement.

In contrast to previous studies, we found evidence of the utility of an affordable and accessible test of CF which low tech, portable, and suitable for low resource settings. Nguyen (2017) suggests that for a game to be fun, it is important to involve rules and competition in pursuit of a goal, and the goal should be challenging, though achievable, to provide a sense of achievement and accomplishment. This provides support that child-friendly assessments do not require expensive stimuli, such as toys or technology, and instead colourful imagery and a game-like format may be sufficient to improve engagement.

Furthermore, this study provides support of the use of simple scoring methods which also improve accessibility of the assessment to clinicians. The measures which provided the most potential were those easiest to calculate, such as summing the number of questions required to reach the target. Furthermore, children's performance was consistent across the four trials, suggesting a single trial of the Alien Game may be sufficient to capture skills, which may further improve accessibility and affordability.

4.5. Critical Evaluation

4.5.1 Strengths

Assessments that have previously been designed specifically for children have focused on expressive executive functions, such as task setting and working memory, and there is limited research on receptive executive functions such as abstraction. Additionally, cognitive assessments often advantage children from Western-European cultures, and there is a need to develop tests which are culturally fair. This study addresses these gaps and provides evidence for the utility of the Alien Game as a child-friendly and more culturally fair test of CF. Additionally, the assessment and scoring procedure were developed to increase affordability and ease to administer.

This study recruited a large sample of children who were diverse in ethnicity, languages spoken, SES, and sex, to be representative of children attending schools in London, which supports the generalisability of the findings. The developments made attempted to improve the Alien Game as a culturally fair assessment of concept formation; according to van de Vijver and Tanzer (2004) cultural fairness of neuropsychological assessments can be improved in several ways. Firstly, the authors propose that construct bias can be avoided by removing culturally specific constructs and language. The Alien Game attempted to meet this criterion by using novel images and reducing language requirements. Secondly, they suggest method bias can be avoided by reducing familiarity with testing materials and stimuli, which may advantage children who have attended school in the West compared with those who have not. They also suggest that method bias can be reduced by limiting the impact of language differences between the examiner and examinee on performance. Language requirements were reduced as much as possible in the Alien Game; participants were permitted to describe the alien attributes, or point to them, if they were unsure of the name. Lastly, they suggest item bias can be improved by assessing individual items for culturally laden constructs. The Alien Game did not include any constructs which may be deemed culturally specific, and included only perceptual categories which are applicable cross culturally.

4.5.2 Limitations

A reliable difference in performance was not found across trial A to trial D for any of the measures suggesting that familiarity with the Alien Game did not lead to improved performance. It is worth noting, however, that most children attending schools in the UK are likely to be familiar with proprietary game ‘Guess Who?’, which may have masked the impact of familiarity with the game format on performance. Different results may have therefore been found if the game had been tested on a sample of children who had grown up in a non-Western culture, and who had not played ‘Guess Who?’ before. To reduce the chance of familiarity with the game ‘Guess Who?’ advantaging some participants over others, and to reduce possible method bias further, the Alien Game could incorporate practice trials to allow children to develop familiarity with the game, and thereby reduce any discrepancy between those who have played ‘Guess Who?’ before and those who have not.

Another limitation of the study is that the sample of children recruited for this study were all from a single primary school in London which reduces the generalisability of the findings. While most children in the classes involved in this study took part, a few were opted out by their parents. The reasons for opting out were not explored, however, this may have resulted in children with certain characteristics or experiences being recruited. For example, it is possible that children with certain diagnoses or difficulties may have been opted out by their parents if it was assumed they would find the task difficult.

Additionally, the school these children were recruited from has small classroom sizes, and the ratio of teachers to students is smaller than the average school in the UK. This may explain why children in this sample demonstrated advanced verbal skills for their age. Performance in the Alien Game, however, was not found to correlate with verbal skills, suggesting that the advanced verbal skills of the sample did not limit the results found, though future research should aim to recruit a sample with more diverse language skills to explore this further. While verbal requirements of the task were limited as much as possible, 20-Questions style tasks do rely on verbal responses. Though the use of perceptual features may support the use of this test with children who have mild language difficulties, it may not be suitable for children who have severe language impairments. Further research is required to investigate the use of this assessment in clinical populations.

Another limitation of this study is that WISC-IV was used rather than the WISC-V and so the norms may overestimate the participant abilities due to the Flynn Effect, which contests that intellectual abilities improve over time (Flynn, 1984). The procedure could also be further developed through developing the task instructions to further clarify the task rules (for example, which question types are permitted), and the stimuli could be improved by removing the alien antennae as some children struggled to name this attribute.

While enjoyability of the Alien Game was assessed, the power imbalance between the researcher and participants may have impacted how able children felt to give a truthful response. Participants were not asked to rate enjoyability of the WISC and so comparisons cannot be made between enjoyability of the Alien Game versus pre-existing measures of CF.

While the Alien Game was designed to be more culturally fair than existing tests, group differences based on English as a first language were also not found for WISC-IV Similarities or Matrix Reasoning, which limits the conclusions which can be drawn regarding the Alien Game as a more culturally fair test than those which currently exist. Furthermore, while this test attempted to produce a culturally fair test of CF, as Cole (1999) suggests, there is no such thing as a culture-free test, since tests of ability are Western constructs. Caution should always be applied when using any cognitive assessment, including the Alien Game, with children who have been brought up in a different culture to that in which the assessment was developed.

4.6. Clinical Implications

Executive functions, including CF, are essential for physical and mental health (Diamond, 2013). CF is also a core skill underlying learning and in the development of general intellectual ability, and it is therefore an important skill for language development and proficiency in many academic subjects including mathematics and physics. Executive functions are also required to support learning behaviour in the classroom such as attention, ability to follow instructions and ignore distractions, as well as behaviour and emotion regulation. Consequently, CF and other executive functions have been directly associated with academic attainment, lifelong achievements, and socioeconomic status in adulthood (Moffitt et al., 2011). For this reason, suitable assessments of concept formation for children are important so difficulties can be identified, and support can be provided.

Difficulties with conceptual skills have been identified in children with neurodevelopmental difficulties such as ASD, ADHD and LD, and children with brain injuries. Diagnostic assessments, however, are often time-consuming and resource heavy (Putra et al., 2020), and so fast but accurate screening of CF difficulties is important. Additionally, current tests that are available may not be providing children with the opportunity to perform their best and thus may under or over-diagnose difficulties. Children require tests appropriate for their stage of cognitive development, such as being engaging and not anxiety-provoking, and utilising simple instructions and stimuli. Such tests should also be suitable for children from a range of cultural backgrounds. This study provided support for the use of the Alien Game to meet these needs.

Socioeconomic status has also been found to impact executive function development (Blair, 2002), possibly as a direct consequence of stress, or the availability of parents and teachers to offer rich learning opportunities (Fitzpatrick et al., 2014). Children from disadvantaged backgrounds are therefore as risk of a vicious cycle where they are at risk of poor academic attainment and adult career attainment compared to more socially advantaged peers. The Alien Game could be used as a screening tool by Educational or Clinical Psychologists to explore conceptual skills if difficulties are suspected before a neuropsychological assessment is considered. This could allow for learning plans to be implemented promptly and may negate the need for a full neuropsychological assessment battery. The game was designed to be easy to administer and score, so it could potentially be administered by teachers at school, however a psychologist may still be required for interpretation of the results. By having affordable and accessible tests of CF for children, difficulties can be identified early, and interventions can be implemented.

4.7. Future Directions

4.7.1 Further Development of the Alien Game Stimuli

Exploration of questions found that participants consistently named the attributes correctly, apart from the antennae which some children struggled to name. These children instead used other descriptors to identify these, such as “tubes,” “things,” or “plants,” or used gestures to communicate their intention to the examiner. Language demands could be further reduced by explicitly stating in the task instructions that pointing is permitted if the name of an attribute is unknown, and/or removing hard to name attributes.

4.7.2 Further Development of the Scoring Procedure

As previously noted, future studies could further develop the task instructions to state whether spatial and compound questions are permitted, and if not, include these as ineffective or unallowed questions. Additionally, the Alien Game Score could be further developed to incorporate abstraction score and total questions, removing the need for the weighted ineffective or unallowed questions score. This could further improve the ease of the scoring procedure.

Task acceptability was measured by asking children to rate how much they enjoyed the game using a Likert-scale. As children were told that the researcher created the game, there was likely an element of participant bias and therefore it is possible that children may not have been completely honest in their responses. Evaluation of acceptability could be explored and whether participant bias can be reduced by collecting feedback anonymously. It would also be useful to collect this same feedback for the WISC-IV subtests so enjoyability can be compared between existing measures of abstraction and the Alien Game.

4.7.3 Reliability

Reliability could be further explored by conducting a test-retest study on a sample of children to explore consistency over time. This may produce practice effects which would need to be taken into consideration. Alternatively, reliability could be explored by comparing future tests involving the Alien Game to the findings of the current study, to explore whether equivalent results are found in different samples.

4.7.4 Generalisability

This study offers further evidence that the Alien Game can be used as a test of abstraction. To improve generalisability of the findings, future studies could test the Alien Game and scoring criteria on a larger sample of children from multiple schools across London or the UK, to capture a more diverse range of backgrounds, ages, and abilities. Efforts should be made to recruit a sample of children with more diversity in verbal skills to further improve generalisability of the findings. Testing the Alien Game on a larger sample will also allow for further development of the performance measures.

Future studies could continue to trial the Alien Game on different age groups to explore whether it can be adapted to be accessible to younger children. This may require development of the task instructions to specify certain rules such as use of pointing to further reduce verbal requirements.

Next, the game could be trialled on a clinical population of children with EF difficulties in a known group validity study, to explore whether the game is sensitive to such difficulties in a clinically useful way. While floor effects were not found in a typically developing sample, it is important to explore whether the Alien Game is accessible for children with cognitive difficulties and whether a range of scores can be found to distinguish degrees of difficulty. This data will also support development of the scoring procedure, as a different pattern of results may emerge in a clinical population.

4.8. Concluding Summary

Overall, this study found evidence to support the Alien Game as a culturally fair measure of nonverbal abstraction abilities in children aged 8 to 11. This task used a game-like procedure and was developed specifically for children to improve engagement and enjoyability. A scoring procedure was developed which showed concurrent validity with an established measure of nonverbal abstraction. Future research could develop the Alien Game further with the aim to establish this as an effective and affordable tool to screen for CF difficulties in children. This will allow for early intervention to support children to develop these important skills.

REFERENCES

Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is it a bird? Is it a plane? Category use in problem-solving in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(5), 555–565. https://doi.org/10.1007/s10803-010-1077-9

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 1–181). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Alt, M., Meyers, C., & Alt, P. M. (2013). Using ratings to gain insight into conceptual development. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1650–1661. https://doi.org/10.1044/1092-4388(2013/11-0317)

Ardila, A., & Moreno, S. (2001). Neuropsychological test performance in Aruaco Indians: An exploratory study. *Journal of the International Neuropsychological Society*, *7*(4), 510–515. https://doi.org/10.1017/S1355617701004076

Audi, R. (2011). *Epistemology: A contemporary introduction to the theory of knowledge* (3rd ed.). Routledge.

Berg, V., Rogers, S. L., McMahon, M., Garrett, M., & Manley, D. (2020). A novel approach to measure executive functions in students: An evaluation of two child-friendly apps. *Frontiers in Psychology*, *11*, 1–15. https://doi.org/10.3389/fpsyg.2020.01702

Bernstein, J. H., & Waber, D. P. (2007). Executive capacities from a developmental perspective. In L. Meltzer, *Executive Function in Education: From Theory to Practice* (pp. 39–54). Guilford Press.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*(3), 180–200. https://doi.org/10.1016/j.dr.2009.05.002

Bhaskar, R. (2010). *Reclaiming reality: A critical introduction to contemporary philosophy*. Routledge. https://doi.org/10.4324/9780203843314

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *The American Psychologist*, *57*(2), 111–127. https://doi.org/10.1037//0003-066x.57.2.111

Blair, C. (2016). Executive function and early childhood education. *Current Opinion in Behavioral Sciences*, *10*, 102–107. https://doi.org/10.1016/j.cobeha.2016.05.009

Blair, C., & Razza, R. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Borke, H. (1975). Piaget's mountains revisited: Changes in the egocentric landscape. *Developmental Psychology*, *11*(2), 240–243. https://doi.org/10.1037/h0076459

Bottino, R. M., Ferlino, L., Ott, M., & Tavella, M. (2007). Developing strategic and reasoning abilities with computer games at primary school level. *Computers & Education*, *49*(4), 1272–1286. https://doi.org/10.1016/j.compedu.2006.02.003

Bracken, B. A. (1986). Incidence of basic concepts in the directions of five commonly used American tests of intelligence. *School Psychology International*, *7*(1), 1–10. https://doi.org/10.1177/014303438600700101

Caillois, R. (2001). *Man, play, and games*. University of Illinois Press.

Catale, C., Meulemans, T., & Thorell, L. B. (2015). The Childhood Executive Function Inventory: Confirmatory factor analyses and cross-cultural clinical validity in a sample of 8-to-11-year-old children. *Journal of Attention Disorders*, *19*(6), 489–495. https://doi.org/10.1177/1087054712470971

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Cattell, R. B., & Cattell, H. E. P. (1973). *Measuring intelligence with the culture fair tests*. Institute for Personality and Ability Testing.

Cole, M. (1999). Culture-free versus culture-based measures of cognition. In R. J. Sternberg, *The nature of cognition* (pp. 645–664). MIT Press.

Condy, E., Kaat, A. J., Becker, L., Sullivan, N., Soorya, L., Berger, N., Berry-Kravis, E., Michalak, C., & Thurm, A. (2021). A novel measure of matching categories for early development: Item creation and pilot feasibility study. *Research in Developmental Disabilities*, *115*, Article: 103993. https://doi.org/10.1016/j.ridd.2021.103993

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423. https://doi.org/10.1016/S1364-6613(03)00197-9

Darlaston-Jones, D. (2007). Making connections: The relationship between epistemology and research methods. *The Australian Community Psychologist*, *19*(1), 19–27.

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis–Kaplan executive function system [D-KEFS].* The Psychological Corporation, Harcourt Brace & Company.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, *333*, 959–964. https://doi.org/10.1126/science.1204529

Dilworth-Bart, J. E. (2012). Does executive function mediate SES and home quality associations with academic readiness? *Early Childhood Research Quarterly*, *27*(3), 416–425. https://doi.org/10.1016/j.ecresq.2012.02.002

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: Moving from correlation to causation. *WIREs Cognitive Science*, *3*(3), 377–386. https://doi.org/10.1002/wcs.1176

Engelmann, J. B., & Pessoa, L. (2014). Motivation sharpens exogenous spatial attention. *Motivation Science*, *7*(3), 64–72. https://doi.org/10.1037/2333-8113.1.S.64

Eslinger, P. J., Grattan, L. M., Damasio, H., & Damasio, A. R. (1992). Developmental consequences of childhood frontal lobe damage. *Archives of Neurology*, *49*(7), 764–769. https://doi.org/10.1001/archneur.1992.00530310112021

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, *1110*(1), 166–174. https://doi.org/10.1016/j.brainres.2006.06.072

Feldman, M. J., & Drasgow, J. (1951). A visual-verbal test for schizophrenia. *Psychiatric Quarterly Supplement*, *25*, 55–64.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Fernandez, A., & Abe, J. (2018). Bias in cross-cultural neuropsychological testing: Problems and possible solutions. *Culture & Brain*, *6*(10), 1–35. https://doi.org/10.1007/s40167-017-0050-2

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., Sheridan, M. A., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, *25*(3), 736–744. https://doi.org/10.1177/0956797613516008

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction*, *30*, 25–31. https://doi.org/10.1016/j.learninstruc.2013.11.003

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Ford, C. B., Kim, H. Y., Brown, L., Aber, J. L., & Sheridan, M. A. (2019). A cognitive assessment tool designed for data collection in the field in low- and middle-income countries. *Research in Comparative and International Education*, *14*(1), 141–157. https://doi.org/10.1177/1745499919829217

Freyberg, P. S. (1966). Concept development in Piagetian terms in relation to school attainment. *Journal of Educational Psychology*, *57*(3), 164–168. https://doi.org/10.1037/h0023411

Gingerich, J. (2018). Freedom and the value of games. *Canadian Journal of Philosophy*, *48*(6), 831–849. https://doi.org/10.1080/00455091.2017.1423224

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/https//doi.org/10.3109/02699052.2014.965210

Göksun, T., George, N. R., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and motion: How young children understand causal events. *Child Development*, *84*(4), 1285–1295. https://doi.org/10.1111/cdev.12035

Gorski, P. S. (2013). "What is critical realism? And why should you care?" *Contemporary Sociology*, *42*(5), 658–670. https://doi.org/10.1177/0094306113499533

Haddlesey, C. (2016). *Exploration of the impact of language and culture on neuropsychological tests* [D.Clin.Psych. thesis, Canterbury Christ Church University]. https://repository.canterbury.ac.uk/download/d88f1d4be2703eae1e94a9d202e587693fa9ed0ad35ad436a034dd51bd73541c/737857/Claire_Haddlesey_MRP_2016.pdf

Heaton, R., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test (WCST) manual revised and expanded*. Psychological Assessment Resources. https//10.1037/t31298-000

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior*. *24*(3), 411–482. https://doi.org/10.1086/504455

Hobson, P., Meara, J., & Taylor, C. (2007). The Weigl Colour-Form Sorting Test: A quick and easily administered bedside screen for dementia and executive dysfunction. *International Journal of Geriatric Psychiatry*, *22*(9), 909–915. https://doi.org/10.1002/gps.1765

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2010). Poor working memory: Impact and interventions. In J. Holmes (Ed.), *Advances in child development and behavior* (1st ed., Vol. 39, pp. 1–43). Academic Press. https://doi.org/10.1016/B978-0-12-374748-8.00001-9

Howard, S. J., & Melhuish, E. (2017). An early years toolbox for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment*, *35*(3), 255–275. https://doi.org/10.1177/0734282916633009

IBM Corp. (2021). *IBM SPSS Statistics for Windows, Version 28.0*. IBM Corp.

Isquith, P. K., Gioia, G., & Espy, K. A. (2004). *Executive function in preschool children: Examination through everyday behaviour*. *26*(1), 403–422. https://doi.org/10.1207/s15326942dn2601_3

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, *20*(3), 573–591. https://doi.org/10.1207/S15326942DN2003_2

Johnson, S. B., Riis, J. L., & Noble, K. G. (2016). State of the art review: Poverty and the developing brain. *Pediatrics*, *137*(4), 1–16. https://doi.org/10.1542/peds.2015-3075

Kagan, J. (1966). A development approach to conceptual growth. In J. J. Klausmeir & C. W. Harris, *Analyses of concept learning* (1st ed., pp. 97–116). Academic Press.

Kim, H. H., Kwon, H. W., Kao, S., Sohn, H. S., Kim, J., Choi, Y., Kim, D., Kang, M., & Park, Y. R. (2020). A study of the possibility of detecting pediatric mild developmental delay through a serious game: A randomized cluster trial in Cambodia. *Studies in Health Technology and Informatics*, *270*(1), 736–740. https://doi.org/10.3233/SHTI200258

Klingberg, T., Vaidya, C. J., Gabrieli, J. D., Moseley, M. E., & Hedehus, M. (1999). Myelination and organization of the frontal white matter in children: A diffusion tensor MRI study. *Neuroreport*, *10*(13), 2817–2821. https://doi.org/10.1097/00001756-199909090-00022.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NESPY: A developmental neuropsychological assessment*. The Psychological Corporation.

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), 1–40. https://doi.org/10.1111/desc.12529

Levine, B., Stuss, D. T., & Milberg, W. P. (1995). Concept generation: Validation of a test of executive functioning in a normal aging population. *Journal of Clinical and Experimental Neuropsychology*, *17*(5), 740–758. https://doi.org/10.1080/01688639508405164

Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*(1–4), 281–297. https://doi.org/10.1080/00207598208247445

Lin, H. L., Lawrence, F. R., & Gorrell, J. (2003). Kindergarten teachers' views of children's readiness for school. *Early Childhood Research Quarterly*, *18*(2), 225–237. https://doi.org/10.1016/S0885-2006(03)00028-0

Lord-Maes, J., & Obrzut, J. E. (1996). Neuropsychological consequences of traumatic brain injury in children and adolescents. *Journal of Learning Disabilities*, *29*(6), 609–617. https://doi.org/10.1177/002221949602900605

Lucenet, J., & Blaye, A. (2014). Age-related changes in the temporal dynamics of executive control: A study in 5- and 6-year-old children. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00831

Luria, A. R. (1979). *The making of mind*. Harvard University Press.

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The Test of Everyday Attention for Children (TEA-CH)*. Thames Valley Test Company.

Matusov, E., & Hayes, R. (2000). Sociocultural critique of Piaget and Vygotsky. *New Ideas in Psychology*, *18*(2), 215–239. https://doi.org/10.1016/S0732-118X(00)00009-X

McCoy, D. C., Zuilkowski, S. S., & Fink, G. (2015). Poverty, physical stature, and cognitive skills: Mechanisms underlying children's school enrollment in Zambia. *Developmental Psychology*, *51*(5), 600–614. https://doi.org/10.1037/a0038924

McIntosh, D. E., Brown, M. L., & Ross, S. L. (1995). Relationship between the Bracken Basic Concept Scale and the differential Ability Scales with an at-risk sample of preschoolers. *Psychological Reports*, *76*(1), 219–224. https://doi.org/10.2466/pr0.1995.76.1.219

McPherson, J., & Burns, N. R. (2008). Assessing the validity of computer-game-like tests of processing speed and working memory. *Behavior Research Methods*, *40*(4), 969–981. https://doi.org/10.3758/BRM.40.4.969

Micheals, S., & Cazden, C. B. (1986). Teacher/child collaboration as oral preparation for literacy. In B. B. Schieffelin & P. Gilmore, *The acquisition of literacy: Ethnographic perspectives* (pp. 132–154). Abblex Pub. Corp.

Microsoft Corporation. (2018). *Microsoft Excel*. Microsoft Corporation.

Minshew, N. J., Goldstein, G., Taylor, H. G., & Siegel, D. J. (1994). Academic achievement in high functioning autistic individuals. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 261–270. https://doi.org/10.1080/01688639408402637

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.1136/bmj.b2535

Mpofu, E., Bracken, B., Van de Vijver, F., & Saklofske, D. (2017). Teaching about intelligence and emotional intelligence. In G. J. Rich, U. P. Gielen, & H. Takooshian, *Internationalizing the taching of psychology* (Eds, pp. 281–296). Information Age Publishing.

National Readership Survey. (2006). *Social Grade*. https://nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice* (pp. xvii, 289). Lawrence Erlbaum Associates Publishers.

Nenty, H. J., & Dinero, T. E. (1981). A cross-cultural analysis of the fairness of the Cattell Culture Fair Intelligence Test using the rasch model. *Applied Psychological Measurement*, *5*(3), 355–368. https://doi.org/10.1177/014662168100500309

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences*, *110*(29), 12138–12143. https://doi.org/10.1073/pnas.1304437110

Nguyen, C. T. (2017). Philosophy of games. *Philosophy Compass*, *12*(8), 1–41. https://doi.org/10.1111/phc3.12426

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, *8*(1), 74–87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

Norman, M. A., Moore, D. J., Taylor, M., Franklin, D., Cysique, L., Ake, C., Lazarretto, D., Vaida, F., & Heaton, R. K. (2011). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-card version. *Journal of Clinical and Experimental Neuropsychology*, *33*(7), 793–804. https://doi.org/10.1080/13803395.2011.559157

Office for National Statistics. (2022). *Ethnic group, England and Wales: Census 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021

Pasnak, R., Hansbarger, A., Dodson, S. L., Hart, J. B., & Blaha, J. (1996). Differential results of instruction at the preoperational/concrete operational transition. *Psychology in the Schools*, *33*(1), 70–83. https://doi.org/10.1002/(SICI)1520-6807(199601)33:1<70::AID-PITS9>3.0.CO;2-#

Pasnak, R., Holt, R., Campbell, J. W., & McCutcheon, L. (1991). Cognitive and achievement gains for kindergartners iostructed in Piagetian Operations. *The Journal of Educational Research*, *85*(1), 5–13. https://doi.org/10.1080/00220671.1991.10702807

Pasnak, R., Savage Greene, M., Ferguson, E. O., & Levit, K. (2006). Applying principles of development to help at-risk preschoolers develop numeracy. *Journal of Psychology*, *140*(2), 155–173. https://doi.org/10.3200/JRLP.140.2.155-173

Pavitt, A. (2017). *Using a game-like procedure as a new test of problem solving and concept formation in children* [Doctoral dissertation, University of East London]. http://uel.da.ulcc.ac.uk/6736/1/Anne Pavitt.pdf

Piaget, J. (1952). *The origins of intelligence in children* (p. 419). International Universities Press. https://doi.org/10.1037/11494-000

Piaget, J. (1995). *Sociological studies* (2nd ed.). Routledge. https://doi.org/10.4324/9780203714065

Pinilla, J., Lopez-Valcarcel, B. G., & Urbanos-Garrido, R. M. (2017). Estimating direct effects of parental occupation on Spaniards' health by birth cohort. *BMC Public Health*, *17*(1), 26. https://doi.org/10.1186/s12889-016-3997-6

Putra, P., Shima, K., & Shimatani, K. (2020). Catchicken: A Serious Game Based on the Go/NoGo Task to Estimate Inattentiveness and Impulsivity Symptoms. *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 152–157. https://doi.org/10.1109/CBMS49503.2020.00036

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, *13*(2), 83–103. https://doi.org/10.1007/BF00992956

Roberts, M. A., Franzen, K. M., Furuseth, A., & Fuller, L. (1995). A developmental study of the Tinker Toy test: Normative and clinical observations. *Applied Neuropsychology*, *2*(3–4), 161–166. https://doi.org/10.1080/09084282.1995.9645355

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. Lloyd, *Cognition and categorization* (pp. 27–48).

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, *52*(3), 326–333. https://doi.org/10.1016/S0278-2626(03)00170-2

Sanghvi, P. (2020). Piaget's theory of cognitive development: A review. *Indian Journal of Mental Health*, *7*(2), 90–96. https://doi.org/10.30877/IJMH.7.2.2020.90-96

Satz, P., & Mogel, S. (1962). An abbreviation of the wais for clinical use. *Journal of Clinical Psychology*, *18*(1), 77–79. https://doi.org/10.1002/1097-4679(196201)18:1<77::AID-JCLP2270180124>3.0.CO;2-R

Schreiber, J. E., Possin, K. L., Girard, J. M., & Rey-Casserly, C. (2014). Executive function in children with attention deficit/hyperactivity disorder: The NIH EXAMINER battery. *Journal of the International Neuropsychological Society*, *20*(1), 41–51. https://doi.org/10.1017/S1355617713001100

Schukajlow, S., & Krug, A. (2014). Are interest and enjoyment important for students' performance? In *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (Vol. 5, pp. 129–136). PME.

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The impact of social disparity on prefrontal function in childhood. *PLoS ONE*, *7*(4), 1–13. https://doi.org/10.1371/journal.pone.0035744

Shulman, C., Yirmiya, N., & Greenbaum, C. W. (1995). From categorization to classification: A comparison among individuals with autism, mental retardation, and normal development. *Journal of Abnormal Psychology*, *104*(4), 601–609. https://doi.org/10.1037/0021-843X.104.4.601

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*(1), 385–401. https://doi.org/10.1207/s15326942dn2601_2

Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Tecoulesco, L., Fein, D., & Naigles, L. R. (2021). What categorical induction variability reveals about typical and atypical development. *Journal of Child Language*, *48*(3), 515–540. https://doi.org/10.1017/S0305000920000392

Teuber, H. L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis*, *32*, 615–656. https://doi.org/1973-08681-001

Thorell, L. B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*, *33*, 536–552. https://doi.org/10.1080/87565640802101516

Torres, C. R. (2018). What counts as part of a game? Reconsidering skills. *Journal of the Philosophy of Sport*, *45*(1), 1–21. https://doi.org/10.1080/00948705.2017.1415150

Tzuriel, D., & Klein, P. S. (1985). The assessment of analogical thinking modifiability among regular, special education, disadvantaged, and mentally retarded children. *Journal of Abnormal Child Psychology*, *13*(4), 539–552. https://doi.org/10.1007/BF00923140

Van de Vijver, F. (1997). Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-Cultural Psychology*, *28*(6), 678–709. https://doi.org/10.1177/0022022197286003

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology / Revue Européenne de Psychologie Appliquée*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wallisch, A., Little, L. M., Dean, E., & Dunn, W. (2018). Executive Function Measures for Children: A Scoping Review of Ecological Validity. *OTJR: Occupation, Participation and Health*, *38*(1), 6–14. https://doi.org/10.1177/1539449217727118

Watson, S. (2016). *The role of executive functions in classroom instruction of students with learning disabilities*. *3*(1), 1–5. https://doi.org/10.4172/2469-9837.1000167

Wechsler, D. (1949). *The Wechsler Intelligence Scale for Children (WISC)*. Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV)*. Pearson.

Wechsler, D. (2014). *The Wechsler intelligence scale for Children—Fifth Edition (WISC-V)*. Pearson.

Weigl, E. (1941). On the psychology of so-called processes of abstraction. *The Journal of Abnormal and Social Psychology*, *36*(1), 3–33. https://doi.org/10.1037/h0055544

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

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: Latent structure. *Developmental Psychology*, *44*, 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Willig, C. (2012). Perspectives on the epistemological bases for qualitative research. In *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 5–21). American Psychological Association. https://doi.org/10.1037/13619-002

Wilson, P. (2004). A Preliminary investigation of an early intervention program: Examining the intervention effectiveness of the Bracken Concept Development Program and the Bracken Basic Concept Scale-Revised with head start students. *Psychology in the Schools*, *41*(3), 301–311. https://doi.org/10.1002/pits.10159

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science*, *16*(2), 90–95. https://doi.org/10.1111/j.0956-7976.2005.00787.x

APPENDICES:

A. Pictures used in Pavitt's (2017) Alien Game

B. Pictures Used in the Alien Game in this Study

C. Participant Information Sheet -Child

D. Participant Consent Form - Child

E. Participant Debrief Form – Child

F. Participant Demographic Record Form

G. Alien Game Response Record Form

H. Likert/visual Analogue Scale as Measure of Task Acceptability

I. CHEXI questionnaire (Thorell & Nyberg, 2008)

J. Study Advertisement

K. Parental / Guardian Information Sheet

L. Parental / Guardian Opt Out Consent Form

M. Parental Debrief Form

N. *Loco Parentis* Consent Form

O. Standardised Instructions for the Alien Game

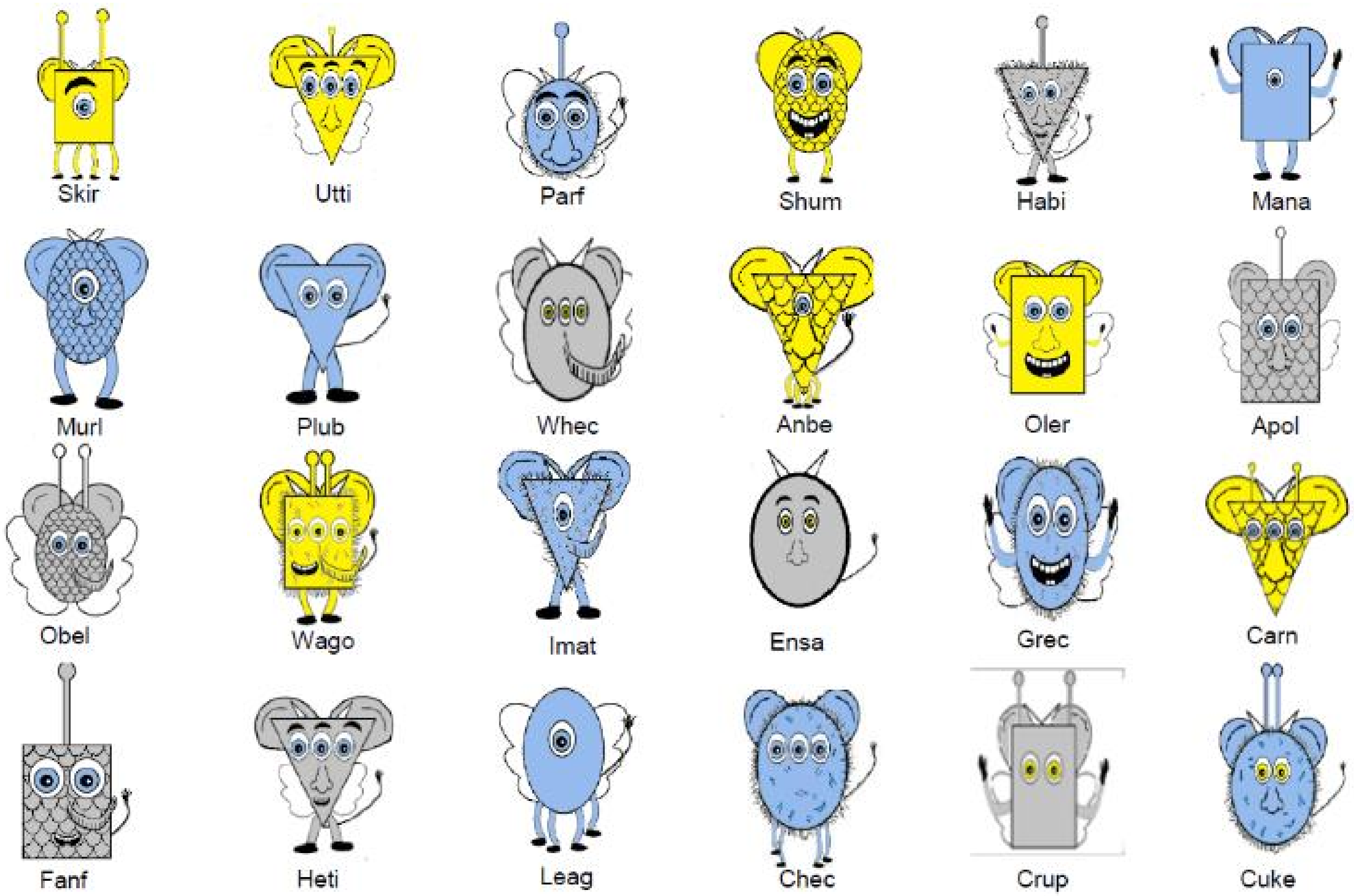
P. Copy of Ethics Application

Q. UEL Risk Assessment Form

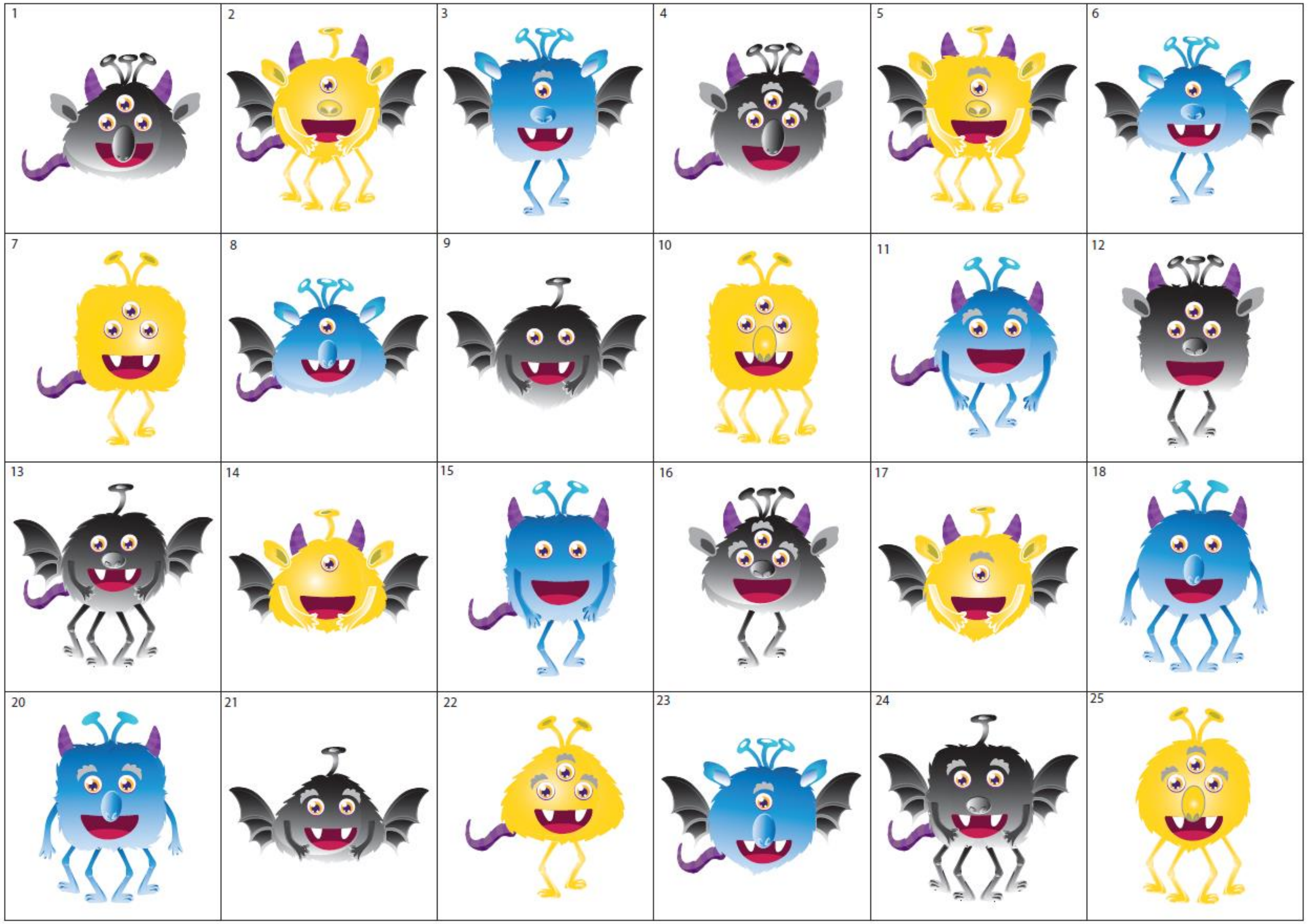
R. Notice of Ethics Review Decision Letter

S. SPSS Output Histograms and Boxplots

APPENDIX A: Pictures used in Pavitt's (2017) Alien Game



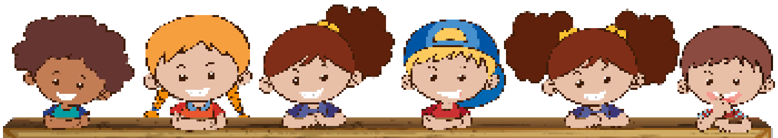
APPENDIX B: Pictures Used in the Alien Game.



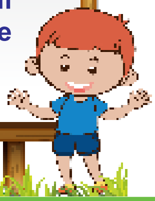
APPENDIX C: Participant Information Sheet -Child

Consent to participate in a research study

Using a Game-Like Procedure as a Test of Executive Functions in Children



You are being invited to take part in a research study. You do not have to take part if you do not want to. Feel free to talk with your family or teachers first before you make your decision. If something does not make sense, or if you have any questions, please talk to us or a teacher.



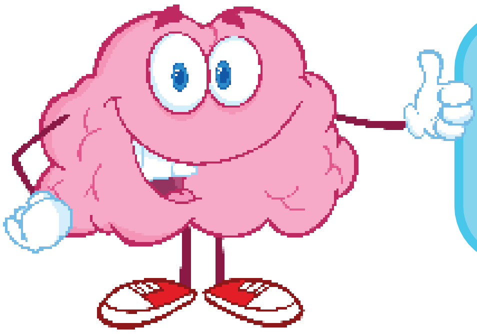
Who are we?

Our names are Alexandros Bardis, Emily Hay and Pinar Marasli. We are all training to be Clinical Psychologists at the University of East London. We are doing some research as part of our studies at university.

Alexandros

Emily

Pinar



What is the purpose of the research?

We have made a game called "The Alien Game" and we want to know if young people your age enjoy this game, and whether it can tell us anything about the way your brain works.

What will you be asked to do?

You will meet with Alex, Emily or Pinar in a quiet room in your school. We will ask you some questions about yourself, like your age and what language you speak at home. You will then be asked to play The Alien Game with one of us. We will then ask you to do some pen and paper tasks that look at your thinking skills. We will also ask your teacher some questions about you, which will help us to assess how good the new game is.

Contact person: Alexandros Bardis, Emily Hay and Pinar Marasli Email: alien.game@uel.ac.uk

Want if you change your mind?
If you decide you do not want to take part anymore, that is fine! You can tell one of us, or you can tell the person who looks after you and they can tell us. You can also change your mind after we have met if it is before January 2023. After January we will have already used your information.

What will happen to your information?
Any information you tell us will be anonymised, which means rather than recording your name we will give you a number, so no one will know it is your information.

The information will be stored in an electronic cloud with a password only we will know. We will look at the information with my supervisor, who we work with. The information will then be put into writing for other psychologists to read.

Who can I contact if I have any questions/concerns?
If you have any questions you can ask the person who looks after you to email us. Our email address is alien.game@uel.ac.uk
They can also contact the research supervisor Dr Matthew Jones Chesters. School of Psychology, University of East London, Water Lane, London E15 4LZ, Email: m.h.jones-chesters@uel.ac.uk.
or
Chair of the School of Psychology Research Ethics Sub-committee: Dr Trishna Patel, School of Psychology, University of East London, Water Lane, London E15 4LZ.
(Email: t.patel@uel.ac.uk)

Contact person: Alexandros Bardis, Emily Hay and Pinar Marasli
Email: alien.game@uel.ac.uk

APPENDIX D: Participant Consent Form - Child

Consent to participate in a research study

USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN

Please read each statement carefully and then tick either yes OR no;

	YES	NO
I have read the information sheet and have been given a copy to keep.		
I have been able to ask questions and have them answered.		
I know that I can change my mind at any time if I don't want to take part anymore without saying why.		
I know that if I no longer want to take part in the study, my answers will not be used.		
I know that I have until the end of January 2023 to change my mind.		
I know that my information and answers will be stored securely and will only be shared with the research team.		
I know what will happen with my information and answers once the research has finished.		
I know that other people will be able to read the final report through the researcher's university.		
I would like to receive a summary of the research once the study has finished and will ask my parent or caregiver to send contact details for this to be sent to.		
I agree to take part in the study.		

Your Name (BLOCK CAPITALS):

Your Signature:

Date:

Researcher's Name (BLOCK CAPITALS):

Researcher's Signature:

Date:

Contact person: Alexandros Bardis, Emily Hay and Pinar Marasli
Email: alien.game@uel.ac.uk

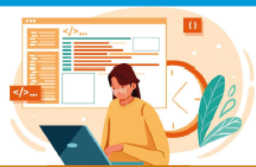
APPENDIX E: Participant Debrief Form - Child

PARTICIPANT DEBRIEF SHEET

USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN

Thank you for taking part in our study!
This study was looking at whether young people your age enjoy The Alien Game, and whether it can tell us anything about the way your brain works.
This document offers information for you now that you have taken part.

How will my data be managed?
Your data will be used by The University of East London. They will make sure your data is held safely. More detailed information is available in the Participant Information Sheet, which you received when you agreed to take part in the research.



What will happen to the results of the research?
We will look at the information with our supervisor, who we work with. The information will then be put into writing for other psychologists to read.
Any information you told us will be anonymised, which means rather than recording your name we have given you a number, so no one will know it is your information.

What if I been negatively affected by taking part?
We do not think you will be negatively affected by taking part in the research, but if you have been affected in any way please talk to your teacher.

Who can I contact if I have any questions/concerns?
If you have any questions you can ask the person who looks after you to email us. Our email address is alien.game@uel.ac.uk
They can also contact the research supervisor Dr Matthew Jones Chesters. School of Psychology, University of East London, Water Lane, London E15 4LZ, Email: m.h.jones-chesters@uel.ac.uk.
or
Chair of the School of Psychology Research Ethics Sub-committee: Dr Trishna Patel, School of Psychology, University of East London, Water Lane, London E15 4LZ.
(Email: t.patel@uel.ac.uk)

Contact person: Alexandros Bardis, Emily Hay and Pinar Marasli
Email: alien.game@uel.ac.uk

APPENDIX F: Participant Demographic Record Form

Demographic Questionnaire

Date of birth: ____

Gender Identity: ____

Ethnicity: ____

Country of birth: ____

Have you always lived in the UK? ____

First language: ____

Main language spoken at home: ____

Parental job title (if known): ____

Education History (for teachers?) (To as teachers: class set? any additional needs? have they attended different schools? All education in England? Any gaps in education?)

____
____
____

Any difficulties with vision or hearing?

____
____

APPENDIX G: Alien Game Response Record Form

Observation Sheet

Date of testing: ___________ Participant ID: ____________

Questions asked by participant:

1. ____
2. ____
3. ____
4. ____
5. ____
6. ____
7. ____
8. ____
9. ____
10. ____

Approach to task (e.g., strategy used?):

____
____
____
____

____

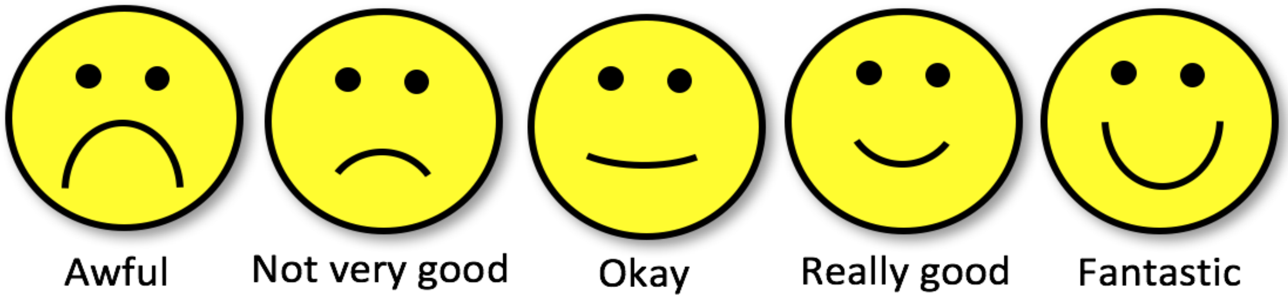
Behavioural observations (e.g., engagement, distractibility, motivation, task enjoyment etc.):

____
____
____
____
____

APPENDIX H: Likert/visual Analogue Scale as Measure of Task Acceptability

How enjoyable was the task today?

Tick the face that shows how you felt:



APPENDIX I: CHEXI questionnaire (Thorell & Nyberg, 2008)

CHILDHOOD EXECUTIVE FUNCTIONING INVENTORY (CHEXI) FOR PARENTS AND EACHERS

Below, you will find a number of statements. Please read each statement carefully and thereafter indicate how well that statement is true for the child. You indicate your response by circling one of the numbers (from 1 to 5) after each statement.

Definitely not true 1	Not true 2	Partially true 3	True 4	Definitely true 5

1. Has difficulty remembering lengthy instructions	1	2	3	4	5
2. Seldom seems to be able to motivate him-/herself to do something that he/she doesn't want to do	1	2	3	4	5
3. Has difficulty remembering what he/she is doing, in the middle of an activity	1	2	3	4	5
4. Has difficulty following through on less appealing tasks unless he/she is promised some type of reward for doing so	1	2	3	4	5
5. Has a tendency to do things without first thinking about what could happen	1	2	3	4	5
6. When asked to do several things, he/she only remembers the first or last	1	2	3	4	5
7. Has difficulty coming up with a different way of solving a problem when he/she gets stuck	1	2	3	4	5
8. When something needs to be done, he/she is often distracted by something more appealing	1	2	3	4	5
9. Easily forgets what he/she is asked to fetch	1	2	3	4	5
10. Gets overly excited when something special is going to happen (e.g., going on a field trip, going to a party)	1	2	3	4	5
11. Has clear difficulties doing things he/she finds boring	1	2	3	4	5
12. Has difficulty planning for an activity (e.g., remembering to bring everything necessary for a field trip or things needed for school)	1	2	3	4	5
13. Has difficulty holding back his/her activity despite being told to do so	1	2	3	4	5
14. Has difficulty carrying out activities that require several steps (e.g., for younger children, getting completely dressed without reminders; for older children, doing all homework independently)	1	2	3	4	5

APPENDIX J: Study Advertisement

University of East London

SCHOOL OF PSYCHOLOGY

THE ALIEN GAME

A new cognitive assessment for children

An exciting opportunity to be part of our Clinical Psychology Doctoral research aiming to create fairer and friendlier cognitive tests for primary-school aged children.

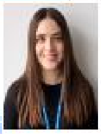
We hope you are interested in this exciting opportunity, and we will contact you again soon to offer the opportunity to discuss in more detail. In the meantime, contact us with any questions.

CONTACT DETAILS:

alien.game@uel.ac.uk

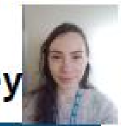
Our Team:

Alex Bardis



Pinar Marasli

Emily Hay



Meabh Foley

PARTICIPANTS

Pupils from Years 1 to Year 6 can participate with parent's consent.

STUDY

The pupil will engage in short, game-like tasks. Teachers and parents may be asked to fill out some short questionnaires.

FREE PPD TRAINING

To thank you for your help, we will offer training to the school team (you can choose from a pre-selected list e.g. how to make CAMHS referrals, and wellbeing strategies).

APPENDIX K: Parental / Guardian Information Sheet

PARTICIPANT INFORMATION SHEET FOR PARENTS

USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN **Contact person:** Alexandros Bardis, Emily Hay or Pinar Marasli

Email: alien.game@uel.ac.uk

Your child is being invited to participate in a research study. Before you decide whether you agree for your child to take part or not, please carefully read through the following information which outlines what their participation would involve. Feel free to talk with others about the study (e.g., friends, family, etc.) before making your decision. If anything is unclear or you have any questions, please do not hesitate to contact us on the above email.

Who are we?

Our names are Alexandros Bardis, Emily Hay and Pinar Marasli and we are Trainee Clinical Psychologists. This study is being conducted as part of our Professional Doctorate in Clinical Psychology at the University of East London. As part of our studies, we are conducting the research that your child is being invited to participate in.

What is the purpose of the research?

We are conducting research into improving neuropsychological tests of executive functions for children. Executive functions are a set of cognitive abilities that includes planning, adjusting, and organising thinking and behaviour.

It is important to understand executive functioning in children, as we use these abilities in everyday life and they impact school attainment, however, most currently available tests of executive functions were designed for adults. They are also often limited by time, cultural norms, and language.

The aim of this study is to assess whether a newly developed game can successfully test executive functions in children in a more engaging and accessible manner than tests that are currently available. This newly developed game hopes to address some of the limitations of existing tests, and if children do find it more engaging it could help us measure these skills more accurately.

Why has your child been invited to take part?

To address the study aims, we are inviting children aged 6-11 to take part in our research.

It is entirely up to you and your child whether your child takes part or not, participation is voluntary.

What will your child be asked to do if I agree for them to take part?

Children will then attend a session with either Alex, Emily or Pinar, where they will be asked some background questions, such as their date of birth, gender identity, ethnicity, country of birth, first language, main language spoken at home and their parents job title (if applicable). They will then be asked to complete some pen and paper neuropsychological measures and a new game developed to measure executive function in children. The session should take about an hour and will take place in a quiet room at their school. We would also ask the child's teacher to fill in a brief questionnaire about the child's ability to plan, adjust and organise their thinking and behaviours in class. The aim of this is to find out whether the measures are related to real-life strengths and/or difficulties.

We will not be able to pay for children's participation in my research, but their participation would be very valuable in helping to develop knowledge and understanding of our research topic.

Can I change my mind?

Yes, you can change your mind at any time and withdraw without explanation, disadvantage, or consequence. If you would like to withdraw your child's data from this study you can do so by letting Alex, Emily or Pinar know via the email address at the top of this letter. If you withdraw, your child’s data will not be used as part of the research.

Separately, you can also request to withdraw your child’s data from being used even after you have taken part in the study, provided that this request is made by the end of January[1] 2022 (after which point the data analysis will begin, and withdrawal will not be possible).

How will the information I provide be kept secure and confidential?

We will anonymously store all data collected on a personal drive, that will be password protected and which only those involved in the research project will have access to. Data will be anonymised through participants being allocated a number which their data will be recorded against; there will be no way of identifying who has been assigned to each number.

For the purposes of data protection, the University of East London is the Data Controller for the personal information processed as part of this research project. The University processes this information under the ‘public task’ condition contained in the General Data Protection Regulation (GDPR). Where the University processes particularly sensitive data (known as ‘special category data’ in the GDPR), it does so because the processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes. The University will ensure that the personal data it processes is held securely and processed in accordance with the GDPR and the Data Protection Act 2018. For more information about how the University processes personal data please see www.uel.ac.uk/about/about-uel/governance/information-assurance/data-protection

What will happen to the results of the research?

Summaries of the data collected will be available in the write-up as a thesis and submitted for assessment. The thesis may be published in an academic journal and will also be publicly accessible on UEL’s online Repository. In all material produced, your child’s identity will remain anonymous, in that, it will not be possible to identify them personally. Some broad demographic information may appear in the thesis and works based on it but that this will not be such as to permit the identification of individual participants.

Anonymised research data will be securely stored by our supervisor, Dr Matthew Jones Chesters, for a maximum of 3 years, following which all data will be deleted.

Who has reviewed the research?

Our research has been approved by the School of Psychology Research Ethics Committee. This means that the Committee's evaluation of this ethics application has been guided by the standards of research ethics set by the British Psychological Society.

Who can I contact if I have any questions/concerns?

If you would like further information about our research or have any questions or concerns, please do not hesitate to contact us. The email address is: alien.game@uel.ac.uk

If you have any questions or concerns about how the research has been conducted, please contact our research supervisor Dr Matthew Jones Chesters, School of Psychology, University of East London, Water Lane, London E15 4LZ,

Email: m.h.jones-chesters@uel.ac.uk.

or

Chair of School Research Ethics Committee: Dr Trishna Patel, School of Psychology, University of East London, Water Lane, London E15 4LZ.

(Email: t.patel@uel.ac.uk)

Thank you for taking the time to read this information sheet

APPENDIX L: Parental / Guardian Opt Out Consent Form

UNIVERSITY OF EAST LONDON

PARENTAL CONSENT OPT-OUT FORM

This form only needs to be returned if you DO NOT want your child to participate

USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN

Your child is being invited to participate in a research study. Before you decide whether you agree for your child to take part or not, please carefully read through the information sheet which outlines what their participation would involve. Feel free to talk with others about the study (e.g., friends, family, etc.) before making your decision. If anything is unclear or you have any questions, please do not hesitate to contact us on alien.game@uel.ac.uk

Your child's participation in the study is voluntary and you can withdraw them at any time before **January 2023**, without explanation or disadvantage. If you withdraw from the study, your child's data will not be used.

Any personal information and data from the research will be securely stored and remain strictly confidential. Only the research team will have access to this information.

Anonymised data may be used in material such as conference presentations, reports, articles in academic journals resulting from the study, though these will not personally identify your child.

If you would like to receive a summary of the research findings once the study has been
completed you can contact the research team via alien.game@uel.ac.uk

If you do not want your child to take part in the study, (1) check the box below, (2) sign the form and date it, and (3) return it to the school within 3 days. You can contact us via alien.game@uel.ac.uk or speak with the school team if you have any questions. Thank you.

Note: If you do not want your child to participate in this study, please complete this form and return to your child's school. You do not need to return this form if you would like for your child to participate.

Child's name (please print)

____

Child's age group____

I have read this form and **do not** grant permission for my child to participate in this study

☐ **No** - My child may **not** take part in this study.

Parent / guardian signature________________________

Date_________________

APPENDIX M: Parental Debrief Form

PARTICIPANT DEBRIEF SHEET

USING A GAME-LIKE PROCEDURE AS A TEST OF EXECUTIVE FUNCTIONS IN CHILDREN

Thank you for your child's participation in our research study into improving neuropsychological tests of executive functions for children. This document offers information that may be relevant in light of them having now taken part.

How will my data be managed?

The University of East London is the Data Controller for the personal information processed as part of this research project. The University will ensure that the personal data it processes is held securely and processed in accordance with the GDPR and the Data Protection Act 2018. More detailed information is available in the Participant Information Sheet, which you received when you agreed to take part in the research.

What will happen to the results of the research?

The research will be written up as a thesis and submitted for assessment. The thesis will be publicly available on UEL's online Repository. Findings will also be disseminated to a range of audiences (e.g., academics, clinicians, public, etc.) through journal articles. In all material produced, your child's identity will remain anonymous, in that, it will not be possible to identify them personally. Some broad demographic information may appear in the thesis and works based on it but that this will not be such as to permit the identification of individual participants.

Anonymised research data will be securely stored by our supervisor, Dr Matthew Jones Chesters for a maximum of 3 years, following which all data will be deleted.

What if I been adversely affected by taking part?

It is not anticipated that your child will have been adversely affected by taking part in the research, and all reasonable steps have been taken to minimise distress or harm of any kind. Nevertheless, it is possible that participation – or its after-effects – may have been challenging, distressing or uncomfortable in some way. If you have any concerns please speak with your child's teacher or school SENCO.

Who can I contact if I have any questions/concerns?

If you would like further information about my research or have any questions or concerns, please do not hesitate to contact us. Our email address is: alien.game@uel.ac.uk

If you have any questions or concerns about how the research has been conducted, please contact our research supervisor, Dr Matthew Jones Chesters, School of Psychology, University of East London, Water Lane, London E15 4LZ,

Email: m.h.jones-chesters@uel.ac.uk.

or

Chair of School Research Ethics Committee: Dr Trishna Patel, School of Psychology, University of East London, Water Lane, London E15 4LZ.

(Email: t.patel@uel.ac.uk)

Thank you for taking part in our study

APPENDIX N: Loco Parentis Consent Form

UNIVERSITY OF EAST LONDON

Using a game-like task as an assessment of concept formation in children
Head Teacher's *Loco Parentis* Form

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 (BLOCK CAPITALS)……………………………………………………………………………………

Name of school (BLOCK CAPITALS): ……………………………………………………………………………………………

Signature of head teacher: ……………………………………………………………………………………………………

Date: …….

Researcher's Name (BLOCK CAPITALS):……………………………………………………………………………………

Researcher's Signature: ………………………………………………………………………………………………………

Date:………

APPENDIX O: Standardised Instructions for the Alien Game

Game procedure

1. Say: *"Today we're going to play a game where you need to guess what alien I am thinking of. You can ask me any question you want that I can reply to with a yes or no answer; the idea is to figure out my alien in as few questions as possible. Remember, you're going to try to figure out what alien I'm thinking of"*
2. START TIMING
3. Every time they ask the wrong type of question, reply "I can only answer yes or no".
4. **If they pause longer than 30 seconds say**: "*Remember you're going to try to figure out what alien I'm thinking of, ask me questions to try to guess the alien I am thinking of, but I can only answer yes or no".*

APPENDIX P: Copy of Ethics Application

UNIVERSITY OF EAST LONDON

School of Psychology

APPLICATION FOR RESEARCH ETHICS APPROVAL

FOR RESEARCH INVOLVING HUMAN PARTICIPANTS

(Updated October 2021)

FOR BSc RESEARCH;

MSc/MA RESEARCH;

PROFESSIONAL DOCTORATE RESEARCH IN CLINICAL, COUNSELLING & EDUCATIONAL PSYCHOLOGY

Section 1 – Guidance on Completing the Application Form (please read carefully)	
1.1	Before completing this application, please familiarise yourself with: ▪ British Psychological Society's Code of Ethics and Conduct ▪ UEL's Code of Practice for Research Ethics ▪ UEL's Research Data Management Policy ▪ UEL's Data Backup Policy
1.2	Email your supervisor the completed application and all attachments as ONE WORD DOCUMENT. Your supervisor will look over your application and provide feedback.
1.3	When your application demonstrates a sound ethical protocol, your supervisor will submit it for review.
1.4	Your supervisor will let you know the outcome of your application. Recruitment and data collection must **NOT** commence until your ethics application has been approved, along with other approvals that may be necessary (see section 7).
1.5	Research in the NHS: ▪ If your research involves patients or service users of the NHS, their relatives or carers, as well as those in receipt of services provided under contract to the NHS, you

	will need to apply for HRA approval/NHS permission (through IRAS). You DO NOT need to apply to the School of Psychology for ethical clearance. ▪ Useful websites: https://www.myresearchproject.org.uk/Signin.aspx https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/hra-approval/ ▪ If recruitment involves NHS staff via the NHS, an application will need to be submitted to the HRA in order to obtain R&D approval. This is in addition to separate approval via the R&D department of the NHS Trust involved in the research. UEL ethical approval will also be required. ▪ HRA/R&D approval is not required for research when NHS employees are not recruited directly through NHS lines of communication (UEL ethical approval is required). This means that NHS staff can participate in research without HRA approval when a student recruits via their own social/professional networks or through a professional body such as the BPS, for example. ▪ The School strongly discourages BSc and MSc/MA students from designing research that requires HRA approval for research involving the NHS, as this can be a very demanding and lengthy process.
1.6	If you require Disclosure Barring Service (DBS) clearance (see section 6), please request a DBS clearance form from the Hub, complete it fully, and return it to applicantchecks@uel.ac.uk. Once the form has been approved, you will be registered with GBG Online Disclosures and a registration email will be sent to you. Guidance for completing the online form is provided on the GBG website: https://fadv.onlinedisclosures.co.uk/Authentication/Login You may also find the following website to be a useful resource: https://www.gov.uk/government/organisations/disclosure-and-barring-service
1.7	Checklist, the following attachments should be included if appropriate: ▪ Study advertisement ▪ Participant Information Sheet (PIS) ▪ Participant Consent Form ▪ Participant Debrief Sheet ▪ Risk Assessment Form/Country-Specific Risk Assessment Form (see section 5) ▪ Permission from an external organisation (see section 7) ▪ Original and/or pre-existing questionnaire(s) and test(s) you intend to use ▪ Interview guide for qualitative studies ▪ Visual material(s) you intend showing participants

Section 2 – Your Details

2.1	**Your name:**	**Emily Hay, Alexandros Bardis, Pinar Marasli**
2.2	**Your supervisor's name:**	**Matthew Jones Chesters**
2.3	**Name(s) of additional UEL supervisors:**	**Emily Hay: Paula Corredor- Lopez. Alexandros Bardis: Trishna Patel. Pinar Marasli: Matthew Boardman** 3rd supervisor (if applicable)

2.4	**Title of your programme:**	**Doctorate in Clinical Psychology**
.5	**UEL assignment submission date:**	**May 22nd 2023**
		Re-sit date (if applicable)

Section 3 – Project Details		
Please give as much detail as necessary for a reviewer to be able to fully understand the nature and purpose of your research.		
3.1	**Study title:** Please note - If your study requires registration, the title inserted here must be the same as that on PhD Manager	Using a Game-Like Procedure as a Test of Executive Functions in Children
3.2	**Summary of study background and aims (using lay language):**	Concept formation is an executive function and can be understood as the ability to identify relationships between objects or events. It is important to understand executive functioning in children, as these abilities have been found to predict school attainment better than IQ (Blair & Razza, 2007). However, most available tests of executive functioning were designed for adults, and are culturally specific. Pavitt (2017) created 'The Alien Game' based on the format of the children's game "Guess Who?" as a more culture fair test of concept formation for children. Pavitt (2017) ran a pilot study to test this approach, and from her results, she identified several areas for improvement. The current study proposes to further develop The Alien Game in the following ways: (a)to improve and refine the materials used in the game; (b) to design a scoring system based on an established measure of concept formation, modified with Pavitt's (2017) suggestions; and (c) to gather richer data on the feasibility of this game as a culturally fair test of concept formation. This study aims to recruit 60-90 children aged 6-11 years. This study will have a cross-sectional correlational design. In addition to playing the game, participants will be asked to complete two existing measures of concept formation, to address concurrent validity. To compare to real-world executive functioning (criterion validity), class teachers will be asked to complete a questionnaire rating (CHEXI) of the child's executive function. We will also address associations between participant demographic data (age, sex and English language facility) and test performance.
3.3	**Research question(s):**	Can a culturally fair test of concept formation be produced that will be engaging to children? Can normative performance characteristics, such as scores and common patterns of responding, which identify normal variation of concept formation be established? Do children engage well with the Alien Game as a measure of concept formation?

		To what extend will the Alien Game scores of 6 – 11 year-olds correlate with other measures of concept formation, e.g. WISC-V Matrix Reasoning and Similarities scores? Will participant demographics play a role in moderating the correlation between Alien Game scores and WISC-V Matrix reasoning and Similarities scores?
3.4	**Research design:**	This study will have a cross-sectional correlational design. Depending on data distributions, parametric or non-parametric procedures (e.g., correlation coefficients, followed up with GLM or regression procedures) will be used to analyse the data and address which variables make unique contributions to test performance. Qualitative data will be used to consider how participants approach the task. Qualitative data will be gathered to understand the quality of the participants responses and strategies, and to determine engagement. Qualitative feedback will be used to determine task enjoyment.
3.5	**Participants:** Include all relevant information including inclusion and exclusion criteria	Participants will be recruited from mainstream primary schools in the London region. As this test aims to be culturally fair, we aim to recruit a sample from a range of backgrounds and abilities. The study will aim to accommodate all needs, and not exclude any participants. Participants will be required to have sufficient English abilities or have an interpreter present to consent to participate. Children with sensory and/or motor function impairments will be included where possible if they volunteer.
3.6	**Recruitment strategy:** Provide as much detail as possible and include a backup plan if relevant	Recruitment of children will be completed through primary schools. Primary schools within London will be contacted via email with details of the study and a poster (see Appendix I) inviting them to take part. A telephone call will be arranged to discuss the details including access to the school and data collection process. We will email the school with all necessary documents and ask them to print information sheets (accessible format for the children) and consent forms for the children and their guardian to read in order to decide whether to take part. Schools will be given the option of using opt-in or opt-out procedure to gain parental consent. Consent will also be gained by the school via the in Loco Parentis form (appendix B). Parents are asked to contact us via email if they have any questions about the study. We will introduce and discuss the study with the child and seek consent or assent as appropriate. Children and parents will be told that they can withdraw their data from the study until the end of the January 2023 if the child / guardian / school change their mind and can stop the study at any point during data collection. Recruitment plan B: To reach out to friends and family who have children within the age range of 6-11, and to recruit via word-of-mouth using the poster (appendix B).
3.7	**Measures, materials or equipment:**	Teachers will be asked to complete the CHEXI as a measure of everyday executive functioning. This is freely available to access online. Two WISC-IV subtests will be administered

	Provide detailed information, e.g., for measures, include scoring instructions, psychometric properties, if freely available, permissions required, etc.	(Similarities and Matrix Reasoning) as single-trial measures of visual and verbal abstraction, to address concurrent validity. These measures will be provided by the supervisor. The participants demographic information (e.g., age, gender identity, ethnicity, country of birth, first language, main language spoken at home and parental job title) will be recorded on a demographics record form produced by the researchers. The Aliens Game will be used to address participants' concept formation abilities. This game will be based on the format of the children's game "Guess Who?" but will consist of a set of cards rather than plastic apparatus. Each card will have a picture of an alien. Each alien will have different characteristics which the participant can ask about in order to identify the target Alien. A record form will be developed to record test performance.	
3.8	**Data collection:** Provide information on how data will be collected from the point of consent to debrief	Parents will be given an information sheet and consent form with the opportunity to opt-out if they do not consent to their child taking part in the study. Participants will be given an information sheet and asked if they consent to taking part and will be given an opportunity to ask questions. The child's teacher will be asked to complete the CHEXI/BRIEF. Before testing begins, demographic data will be collected from the participant (see Appendix B). The Aliens Game will then be administered, beginning with a training trial consisting of 4-6 cards and feedback. The game will then be administered, and it is expected to last around 15 minutes. Testing will take place in a quiet private room within the school, and children will be given breaks between tasks. Following administration of the game the WISC-IV Matrix Reasoning and Similarities subtests will be administered. Participant feedback will then be sought to determine engagement and enjoyment. Overall, we expect the testing procedure to last 45 minutes per child.	
3.9	**Will you be engaging in deception?**	**YES** ☐	**NO** ☒
	If yes, what will participants be told about the nature of the research, and how/when will you inform them about its real nature?	If you selected yes, please provide more information here	
3.10	**Will participants be reimbursed?**	**YES** ☐	**NO** ☒
	If yes, please detail why it is necessary.	If you selected yes, please provide more information here	
	How much will you offer? Please note - This must be in the form of vouchers, not cash.	Please state the value of vouchers	
3.11	**Data analysis:**	This study will use multiple regressions to analyse which variables make a unique contribution to test performance. Therefore, demographic data such as age, gender identity, ethnicity, and first language will act as independent variables and performance on the Alien Game will act at the dependent variable. Scores on WISC-IV Matrix Reasoning and Similarities tests will be compared to performance on the Alien Game to establish concurrent validity, and teacher ratings on	

		the CHEXI/BRIEF will measure predictive validity to real-world executive functioning. Qualitative data will be used to consider how participants approach the task.

Section 4 – Confidentiality, Security and Data Retention

It is vital that data are handled carefully, particularly the details about participants. For information in this area, please see the UEL guidance on data protection, and also the UK government guide to data protection regulations.

If a Research Data Management Plan (RDMP) has been completed and reviewed, information from this document can be inserted here.

4.1	**Will the participants be anonymised at source?**	**YES** ☐	**NO** ☒
	If yes, please provide details of how the data will be anonymised.		
4.2	**Are participants' responses anonymised or are an anonymised sample?**	**YES** X	**NO** ☐
	If yes, please provide details of how data will be anonymised (e.g., all identifying information will be removed during transcription, pseudonyms used, etc.).	Participant's data will be pseudonymised by allocating to each participant a code to corresponding their data. The participant code will be used instead of names in the database. Participant names and codes will be stored in a separate password-protected file. All data, including identifying information will be securely stored in password-protected files in accordance with GDPR regulations. At the end of the study participant names and associated codes will be destroyed. The remaining data will be help for up to two years to support publication of the results.	

4.3	**How will you ensure participant details will be kept confidential?**	Any information which is not anonymous e.g., consent forms, will be scanned and stored securely, then deleted once the research has been completed and assessed. All data will be pseudonymised through recording against an allocated number.
4.4	**How will data be securely stored and backed up during the research?** Please include details of how you will manage access, sharing and security	Folders or documents containing data will be password protected and stored securely on UEL One Drive.
4.5	**Who will have access to the data and in what form?** (e.g., raw data, anonymised data)	The only person who will have access to the data are those named in this application and the Director of Studies; it is possible that access to the data may be requested by thesis examiners.
4.6	**Which data are of long-term value and will be retained?** (e.g., anonymised interview transcripts, anonymised databases)	Anonymised database of quantitative data will be retained for three years.
4.7	**What is the long-term retention plan for this data?**	The data will be kept for three years following the completion of the research. Following submission of the thesis, data will be retained by the Director of Studies and deleted after three years.

4.8	**Will anonymised data be made available for use in future research by other researchers?**	**YES** ☐	**NO** ☒
	If yes, have participants been informed of this?	**YES** ☐	**NO** ☐
4.9	**Will personal contact details be retained to contact participants in the future for other research studies?**	**YES** ☐	**NO** ☒
	If yes, have participants been informed of this?	**YES** ☐	**NO** ☐

Section 5 – Risk Assessment

If you have serious concerns about the safety of a participant, or others, during the course of your research please speak with your supervisor as soon as possible. If there is any unexpected occurrence while you are collecting your data (e.g., a participant or the researcher injures themselves), please report this to your supervisor as soon as possible.			
5.1	**Are there any potential physical or psychological risks to participants related to taking part?** (e.g., potential adverse effects, pain, discomfort, emotional distress, intrusion, etc.)	**YES** ☒	**NO** ☐
	If yes, what are these, and how will they be minimised?	There is a risk of taking part in any in-person research during this endemic phase of the COVID19 pandemic. To minimise risk of infection for the participant, current guidelines will be followed i.e. masks will be worn, the room will be large enough for social distancing and hands and surfaces will be regularly washed/sanitized. The researchers will be completing lateral flow tests twice a week and will isolate for 10 days if the test is positive. Public transport will be avoided where possible when travelling, if this is not possible, the safest routes will be taken. The researchers will adhere to the school's process for risk assessments	
5.2	**Are there any potential physical or psychological risks to you as a researcher?**	**YES** ☒	**NO** ☐
	If yes, what are these, and how will they be minimised?	There is a small risk of completing the research during this endemic phase of the pandemic. To minimise risk of infection for the researcher, guidelines will be followed i.e., masks will be worn, the room will be large enough for social distancing and hands and surfaces will be regularly washed/sanitized. The researchers have received both doses of the vaccine and will be completing lateral flow tests twice a week. Public transport will be avoided where possible when travelling, if this is not possible, the safest routes will be taken. The researcher will adhere to the school's process for risk assessments.	

<table>
<tr><td>5.3</td><td>If you answered yes to either 5.1 and/or 5.2, you will need to complete and include a General Risk Assessment (GRA) form (signed by your supervisor). Please confirm that you have attached a GRA form as an appendix:</td><td colspan="3">YES
☒</td></tr>
<tr><td>5.4</td><td>If necessary, have appropriate support services been identified in material provided to participants?</td><td>YES
☐</td><td>NO
☐</td><td>N/A
☒</td></tr>
<tr><td rowspan="2">5.5</td><td>Does the research take place outside the UEL campus?</td><td>YES
☒</td><td colspan="2">NO
☐</td></tr>
<tr><td>If yes, where?</td><td colspan="3">The data collection will take place on primary school campuses.</td></tr>
<tr><td rowspan="3">5.6</td><td>Does the research take place outside the UK?</td><td>YES
☐</td><td colspan="2">NO
☒</td></tr>
<tr><td>If yes, where?</td><td colspan="3">Please state the country and other relevant details</td></tr>
<tr><td>If yes, in addition to the General Risk Assessment form, a Country-Specific Risk Assessment form must also be completed and included (available in the Ethics folder in the Psychology Noticeboard). Please confirm a Country-Specific Risk Assessment form has been attached as an appendix.
Please note - A Country-Specific Risk Assessment form is not needed if the research is online only (e.g., Qualtrics survey), regardless of the location of the researcher or the participants.</td><td colspan="3">YES
☐</td></tr>
<tr><td>5.7</td><td colspan="4">Additional guidance:
▪ For assistance in completing the risk assessment, please use the AIG Travel Guard website to ascertain risk levels. Click on ‘sign in’ and then ‘register here’ using policy # 0015865161. Please also consult the Foreign Office travel advice website for further guidance.
▪ For on campus students, once the ethics application has been approved by a reviewer, all risk assessments for research abroad must then be signed by the Director of Impact and Innovation, Professor Ian Tucker (who may escalate it up to the Vice Chancellor).
▪ For distance learning students conducting research abroad in the country where they currently reside, a risk assessment must also be carried out. To minimise risk, it is recommended that such students only conduct data collection online. If the project is deemed low risk, then it is not necessary for the risk assessment to be signed by the Director of Impact and Innovation. However, if not deemed low risk, it must be signed by the Director of Impact and Innovation (or potentially the Vice Chancellor).</td></tr>
</table>

	▪ Undergraduate and M-level students are not explicitly prohibited from conducting research abroad. However, it is discouraged because of the inexperience of the students and the time constraints they have to complete their degree.

Section 6 – Disclosure and Barring Service (DBS) Clearance			
6.1	**Does your research involve working with children (aged 16 or under) or vulnerable adults (*see below for definition)?** If yes, you will require Disclosure Barring Service (DBS) or equivalent (for those residing in countries outside of the UK) clearance to conduct the research project	**YES** ☒	**NO** ☐
	* You are required to have DBS or equivalent clearance if your participant group involves: (1) Children and young people who are 16 years of age or under, or (2) 'Vulnerable' people aged 16 and over with particular psychiatric diagnoses, cognitive difficulties, receiving domestic care, in nursing homes, in palliative care, living in institutions or sheltered accommodation, or involved in the criminal justice system, for example. Vulnerable people are understood to be persons who are not necessarily able to freely consent to participating in your research, or who may find it difficult to withhold consent. If in doubt about the extent of the vulnerability of your intended participant group, speak with your supervisor. Methods that maximise the understanding and ability of vulnerable people to give consent should be used whenever possible.		
6.2	**Do you have DBS or equivalent (for those residing in countries outside of the UK) clearance to conduct the research project?**	**YES** X	**NO** ☐
6.3	**Is your DBS or equivalent (for those residing in countries outside of the UK) clearance valid for the duration of the research project?**	**YES** X	**NO** ☐
6.4	**If you have current DBS clearance, please provide your DBS certificate number:**	Emily Hay: 001784322516; Alexandros Bardis: 001584640901; Pinar Marasli: 001687764808	
	If residing outside of the UK, please detail the type of clearance and/or provide certificate number.	Please provide details of the type of clearance, including any identification information such as a certificate number	
6.5	**Additional guidance:** ▪ If participants are aged 16 or under, you will need two separate information sheets, consent forms, and debrief forms (one for the participant, and one for their parent/guardian). ▪ For younger participants, their information sheets, consent form, and debrief form need to be written in age-appropriate language.		

Section 7 – Other Permissions			
7.1	**Does the research involve other organisations (e.g., a school, charity, workplace, local authority, care home, etc.)?**	**YES** ☒	**NO** ☐
	If yes, please provide their details.	Schools will be recruited once ethical approval has been granted for the research to take place.	
	If yes, written permission is needed from such organisations (i.e., if they are helping you with recruitment and/or data collection, if you are collecting data on their premises, or if you are using any material owned by the institution/organisation). Please confirm that you have attached written permission as an appendix.	**YES** ☐	
7.2	**Additional guidance:** ▪ Before the research commences, once your ethics application has been approved, please ensure that you provide the organisation with a copy of the final, approved ethics application or approval letter. Please then prepare a version of the consent form for the organisation themselves to sign. You can adapt it by replacing words such as 'my' or 'I' with 'our organisation' or with the title of the organisation. This organisational consent form must be signed before the research can commence. ▪ If the organisation has their own ethics committee and review process, a SREC application and approval is still required. Ethics approval from SREC can be gained before approval from another research ethics committee is obtained. However, recruitment and data collection are NOT to commence until your research has been approved by the School and other ethics committee/s.		

Section 8 – Declarations		
8.1	**Declaration by student. I confirm that I have discussed the ethics and feasibility of this research proposal with my supervisor:**	**YES** ☒
8.2	**Student's name:** (Typed name acts as a signature)	Alexandros Bardis, Emily Hay and Pinar Marasli
8.3	**Student's number:**	**U2075206; U2075197; U2075213**
8.4	**Date:**	**20/06/2022**
Supervisor's declaration of support is given upon their electronic submission of the application		

APPENDIX Q: UEL Risk Assessment Form

UEL Risk Assessment Form

Name of Assessor:	Alexandros Bardis, Emily Hay and Pinar Marasli	**Date of Assessment:**	16/05/2022
Activity title:	Thesis Recruitment	**Location of activity:**	UEL Campuses at Docklands, Stratford and Primary schools that we recruit to take part following ethical approval of the study
Signed off by Manager: (Print Name)	Matthew Jones Chesters	**Date and time: (if applicable)**	Summer and Autumn Term (Between June 2022 and March 2023)

Please describe the activity/event in as much detail as possible (include nature of activity, estimated number of participants, etc.).
If the activity to be assessed is part of a fieldtrip or event please add an overview of this below:

Research project as part of Professional Doctorate in Clinical Psychology. Participants will be sat in a quiet room at their school with the researcher present. They will be asked to complete a number of questionnaires and pen and paper tasks. Participation will last about 1 hour. We aim to recruit 20-30 children.

For the completion of our research project/thesis we plan to go into schools to recruit participants. The population is children aged 6 to 11. We aim to recruit 60-90 children. We are currently liaising with schools to gain permission to come in and decide how and when this will be done, depending on the restrictions in place. If it is safe to do so, we plan to complete our recruitment in the Autumn And Winter school term, from October 2022 to January 2023. We hope to assess 4 young people in a day each, which means we will need to spend around 8 days in the school, which will be spread out across the two terms, depending on the school's availability. We plan to meet with each young person individually to complete a battery of neuropsychological assessments, a newly developed game assessing executive functioning and a demographic questionnaire. Teachers will also be asked to complete a questionnaire about the participants behaviour. We plan to complete each session with a young person within an hour. When in the school and meeting with the young person, will wear a mask at all times and regularly wash and sanitise our hands and any equipment. If possible, we will also request a room with ventilation and the ability to social distance from one another. The resources we will be using are neuropsychological tests, questionnaires and the newly developed game, all of which will be provided by ourselves. We will also provide the school with a copy of our DBS certificates.

Overview of FIELD TRIP or EVENT:

As above

Guide to risk ratings:

a) Likelihood of Risk	b) Hazard Severity	c) Risk Rating (a x b = c)
1 = Low (Unlikely)	1 = Slight (Minor / less than 3 days off work)	1-2 = Minor (No further action required)
2 = Moderate (Quite likely)	2= Serious (Over 3 days off work)	3-4 = Medium (May require further control measures)
3 = High (Very likely or certain)	3 = Major (Over 7 days off work, specified injury or death)	6/9 = High (Further control measures essential)

Hazards attached to the activity

Hazards identified	Who is at risk?	Existing Controls	Likelihood	Severity	Residual Risk Rating (Likelihood x Severity)	Additional control measures required (if any)	Final risk rating
Obstruction of safe exit routes in event of fire or other emergency, due to blocking of doors/thoroughfare/ fire exit routes with tables, chairs or banners.	Staff Students Researcher	On day, researchers will make sure they are aware of where the fire exits are in relation to the location/room used and make sure tables and chairs do not obstruct exits/entrances or routes.	1	2	2	Ensure placement of objects is monitored throughout the day.	2

Slip or trip hazard due to promotional literature or freebies, or rubbish, being dropped on the floor.	Staff, Students, Researcher	Be vigilant on the day to make sure that belongings do not get left on the floor, ensuring anything that is dropped is picked up immediately and ensuring electrical equipment, such as a laptop charger, is in an appropriate place and not a trip hazard. Ensuring bins and cleaning equipment such as paper towels are available.	2	1	2	Ensure this is monitored throughout the day.	2
Infection of covid-19	Ourselves and students whom participate	Wearing a face mask at all times, social distancing where possible, being in a ventilated room, washing and sanitising hands and equipment regularly. Any students who display symptoms or test positive for covid will not participate for at least 14 days, likewise, if a researcher displays symptoms or test positive for covid, that individual will not visit schools for at least 14 days. we also have received both doses of our covid-19 vaccine and booster.	2	2	4		4

Review Date

APPENDIX R: Notice of Ethics Review Decision Letter

University of East London

School of Psychology Ethics Committee

NOTICE OF ETHICS REVIEW DECISION LETTER

For research involving human participants

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

Reviewer: Please complete sections in **blue** | **Student:** Please complete/read sections in orange

Details	
Reviewer:	**Fiorentina Sterkaj**
Supervisor:	**Matthew Jones Chesters**
Student:	**Emily Hay, Alexandros Bardis, Pinar Marasli**
Course:	**Prof Doc Clinical Psychology**
Title of proposed study:	Please type title of proposed study

Checklist (Optional)			
	YES	**NO**	**N/A**
Concerns regarding study aims (e.g., ethically/morally questionable, unsuitable topic area for level of study, etc.)	☐	☐	☐
Detailed account of participants, including inclusion and exclusion criteria	☐	☐	☐
Concerns regarding participants/target sample	☐	☐	☐
Detailed account of recruitment strategy	☐	☐	☐
Concerns regarding recruitment strategy	☐	☐	☐
All relevant study materials attached (e.g., freely available questionnaires, interview schedules, tests, etc.)	☐	☐	☐
Study materials (e.g., questionnaires, tests, etc.) are appropriate for target sample	☐	☐	☐
Clear and detailed outline of data collection	☐	☐	☐
Data collection appropriate for target sample	☐	☐	☐

If deception being used, rationale provided, and appropriate steps followed to communicate study aims at a later point	☐	☐	☐
If data collection is not anonymous, appropriate steps taken at later stages to ensure participant anonymity (e.g., data analysis, dissemination, etc.) – anonymisation, pseudonymisation	☐	☐	☐
Concerns regarding data storage (e.g., location, type of data, etc.)	☐	☐	☐
Concerns regarding data sharing (e.g., who will have access and how)	☐	☐	☐
Concerns regarding data retention (e.g., unspecified length of time, unclear why data will be retained/who will have access/where stored)	☐	☐	☐
If required, General Risk Assessment form attached	☐	☐	☐
Any physical/psychological risks/burdens to participants have been sufficiently considered and appropriate attempts will be made to minimise	☐	☐	☐
Any physical/psychological risks to the researcher have been sufficiently considered and appropriate attempts will be made to minimise	☐	☐	☐
If required, Country-Specific Risk Assessment form attached	☐	☐	☐
If required, a DBS or equivalent certificate number/information provided	☐	☐	☐
If required, permissions from recruiting organisations attached (e.g., school, charity organisation, etc.)	☐	☐	☐
All relevant information included in the participant information sheet (PIS)	☐	☐	☐
Information in the PIS is study specific	☐	☐	☐
Language used in the PIS is appropriate for the target audience	☐	☐	☐
All issues specific to the study are covered in the consent form	☐	☐	☐
Language used in the consent form is appropriate for the target audience	☐	☐	☐
All necessary information included in the participant debrief sheet	☐	☐	☐
Language used in the debrief sheet is appropriate for the target audience	☐	☐	☐
Study advertisement included	☐	☐	☐
Content of study advertisement is appropriate (e.g., researcher's personal contact details are not shared, appropriate language/visual material used, etc.)	☐	☐	☐

Decision options	
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.
APPROVED - BUT MINOR AMENDMENTS ARE REQUIRED <u>BEFORE</u> THE RESEARCH COMMENCES	In this circumstance, the student must confirm with their supervisor that all minor amendments have been made **<u>before</u>** the research commences. Students are to do this by filling in the confirmation box at the end of this form once all amendments have been attended to and emailing a copy of this decision notice to the supervisor. The supervisor will then forward the student's confirmation to the School for its records.

	Minor amendments guidance: typically involve clarifying/amending information presented to participants (e.g., in the PIS, instructions), further detailing of how data will be securely handled/stored, and/or ensuring consistency in information presented across materials.
NOT APPROVED - MAJOR AMENDMENTS AND RE-SUBMISSION REQUIRED	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. **Major amendments guidance:** typically insufficient information has been provided, insufficient consideration given to several key aspects, there are serious concerns regarding any aspect of the project, and/or serious concerns in the candidate's ability to ethically, safely and sensitively execute the study.

Decision on the above-named proposed research study	
Please indicate the decision:	Please select your decision

Minor amendments
Please clearly detail the amendments the student is required to make
Section 3.3 rephrase research question to reflect a more decisive investigative approach **Section 3.6 Provide more detail re your recruitment strategy, how will you decide which schools to approach, how will you gain access to the school. What is the backup plan if that does not work? What if School/s approve but parents are not willing to allow their children to participate.** **Appendix D. This can be less wordy and further simplified for the participants**

Major amendments
Please clearly detail the amendments the student is required to make

Assessment of risk to researcher		
Has an adequate risk assessment been offered in the application form?	YES ☒	NO ☐
	If no, please request resubmission with an adequate risk assessment.	
If the proposed research could expose the researcher to any kind of emotional, physical or health and safety hazard, please rate the degree of risk:		
HIGH	Please do not approve a high-risk application. Travel to countries/provinces/areas deemed to be high risk should not be permitted and an application not be approved on this basis. If unsure, please refer to the Chair of Ethics.	☐
MEDIUM	Approve but include appropriate recommendations in the below box.	☐
LOW	Approve and if necessary, include any recommendations in the below box.	☒
Reviewer recommendations in relation to risk (if any):	Please insert any recommendations	

Reviewer's signature	
Reviewer: (Typed name to act as signature)	Dr Fiorentina Sterkaj
Date:	27/10/2022
This reviewer has assessed the ethics application for the named research study on behalf of the School of Psychology Ethics Committee	
RESEARCHER PLEASE NOTE For the researcher and participants involved in the above-named study to be covered by UEL's Insurance, prior ethics approval from the School of Psychology (acting on behalf of the UEL Ethics Committee), and	

APPENDIX S: SPSS Output Histograms and Boxplots

Figure S1

Boxplot of Participant Performance in Total Questions Measure

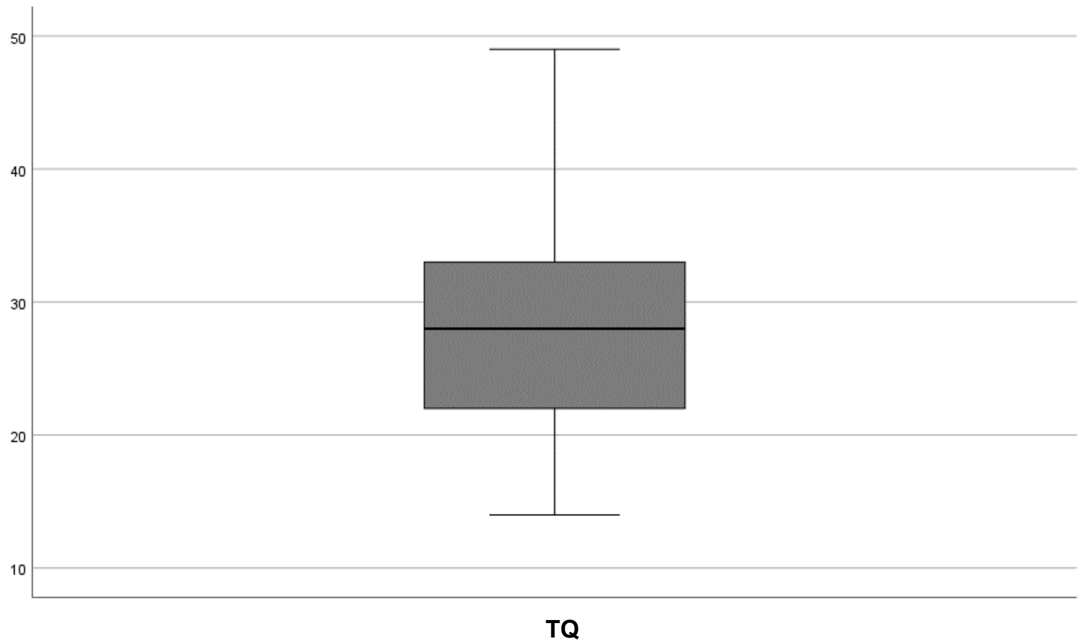


Figure S2

H

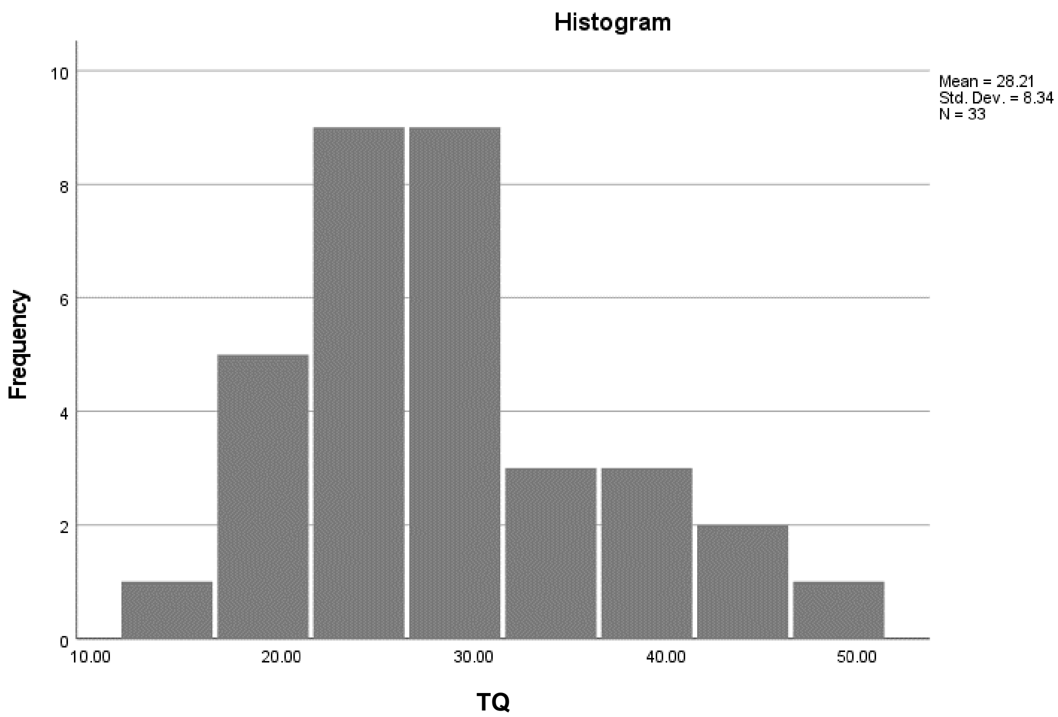


Figure S3

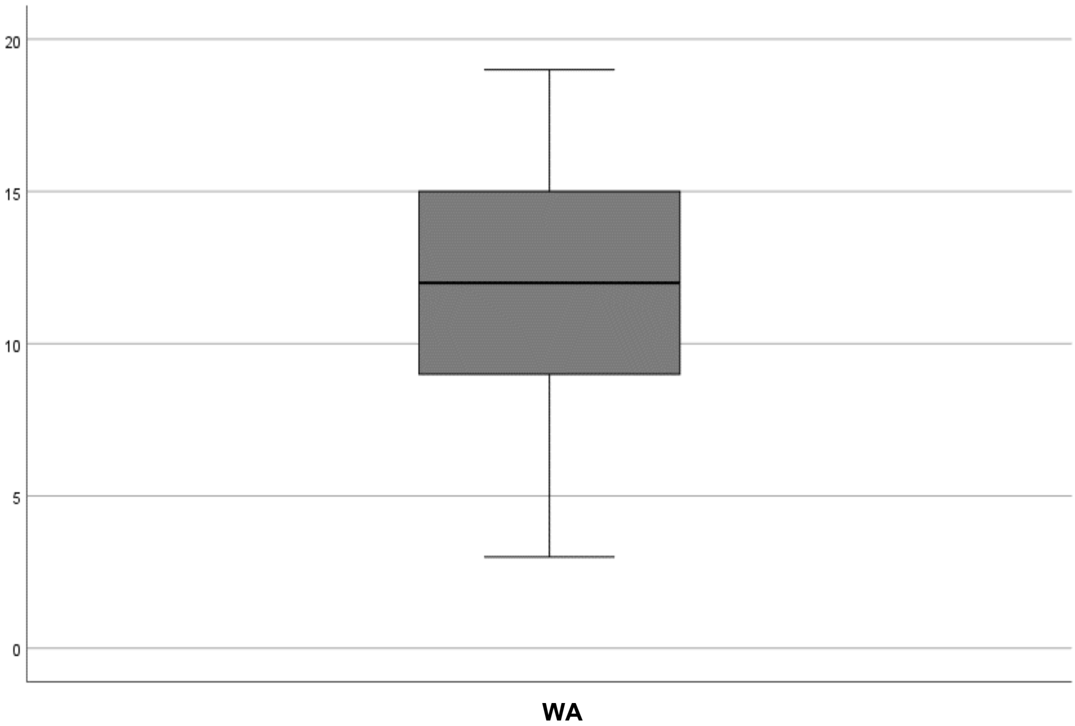


Figure S4

Histogram of Participant Performance in Weighted Achievement Measure

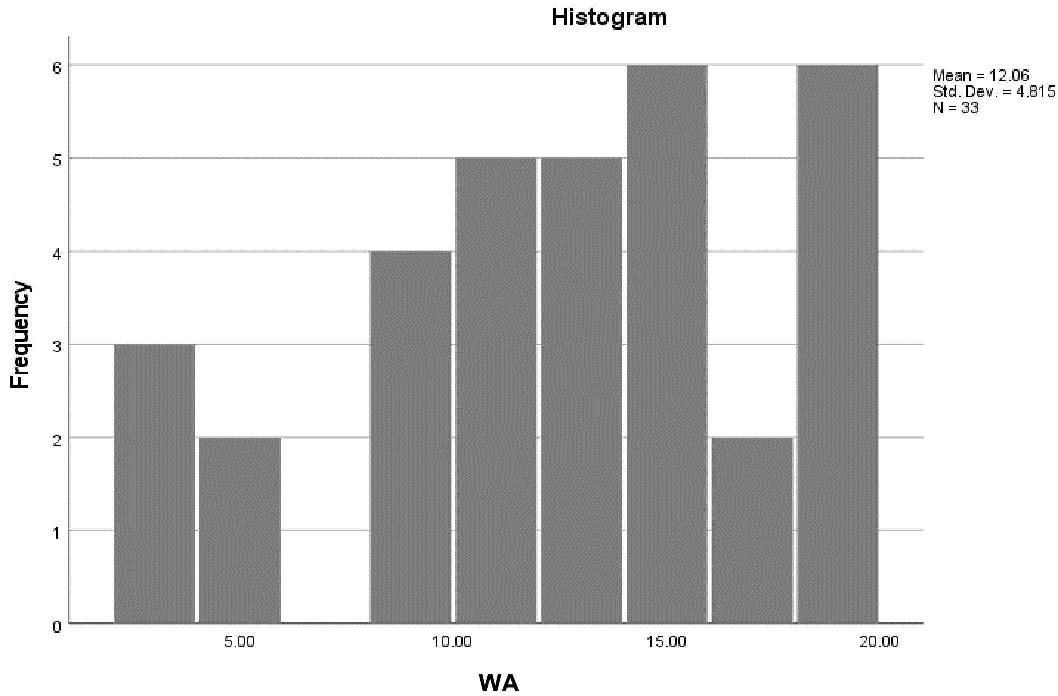


Figure S5

Boxplot of Participant Performance in Initial Abstraction Measure

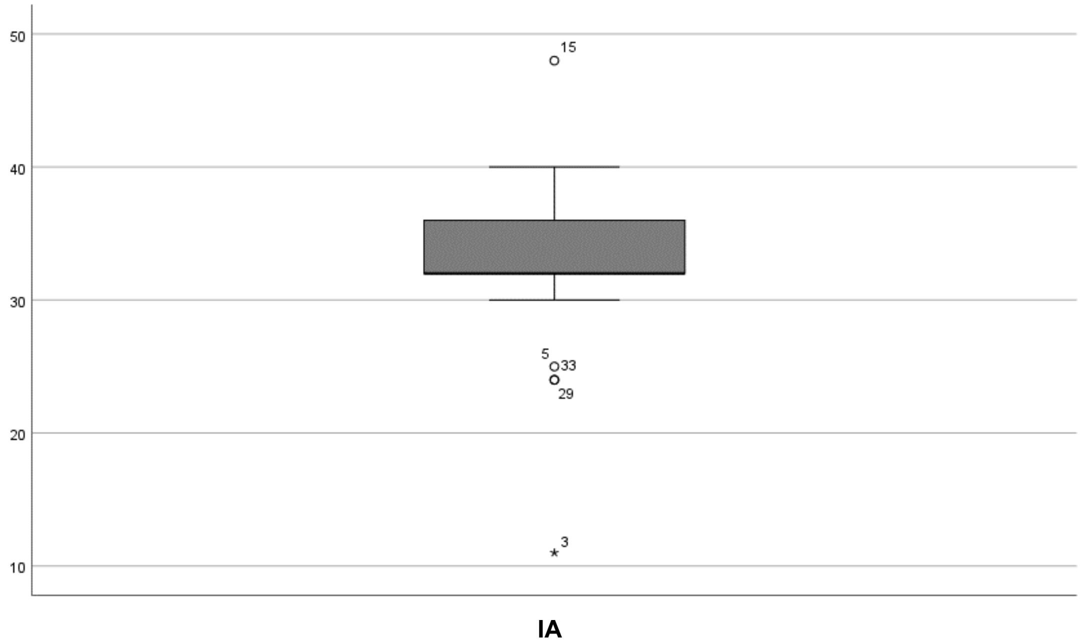


Figure S6

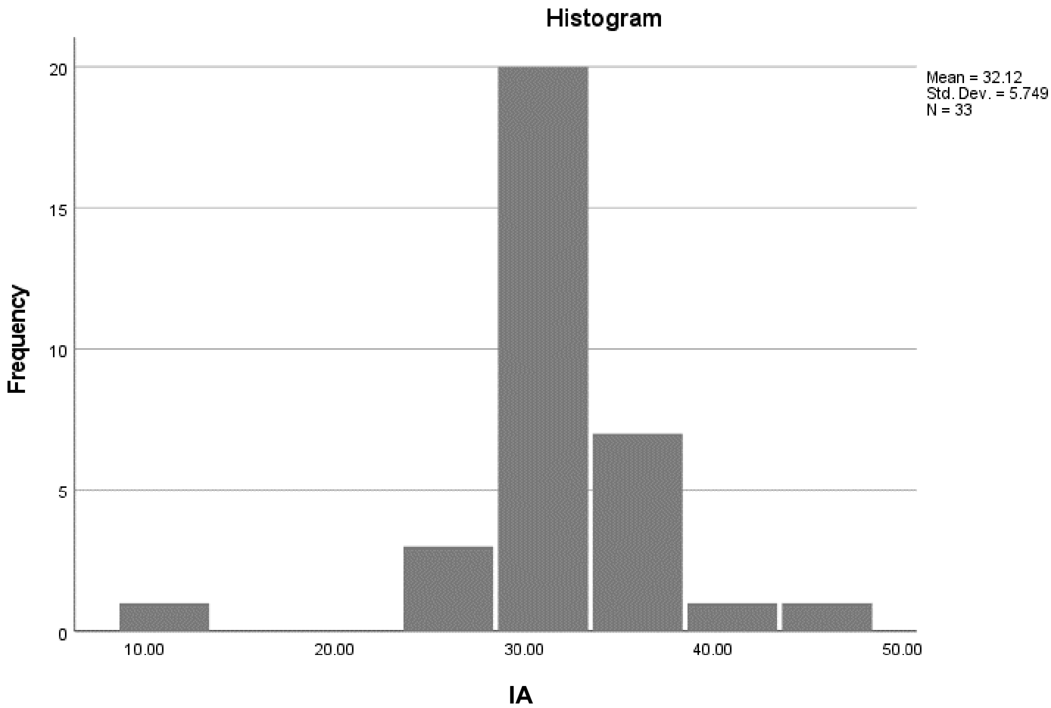


Figure S7

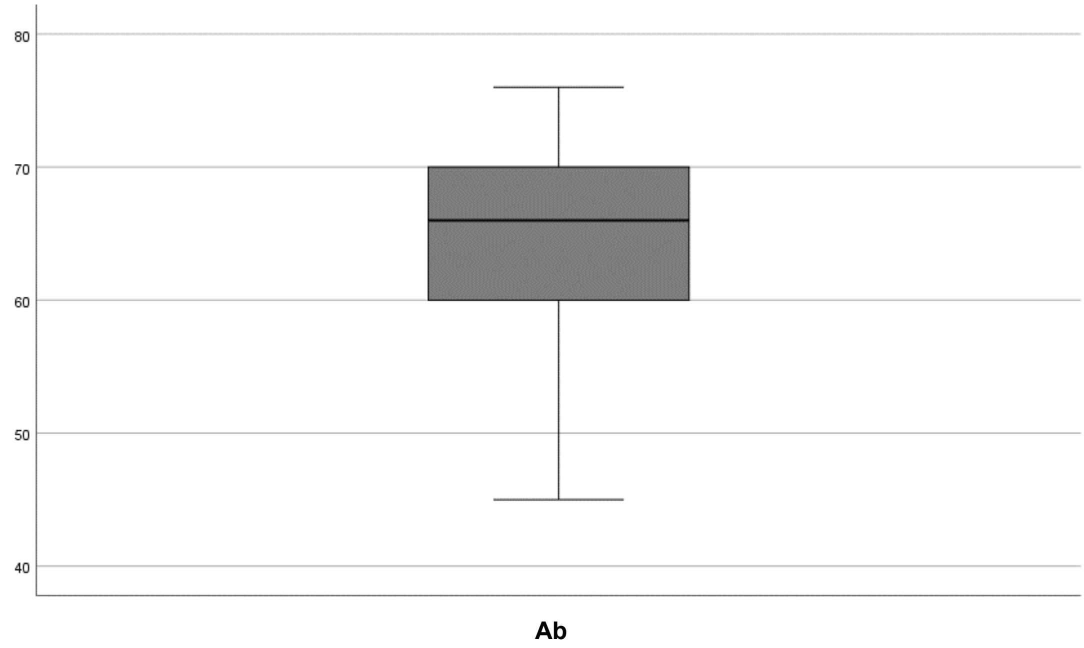


Figure S8

Histogram of Participant Performance in Abstraction Score Measure

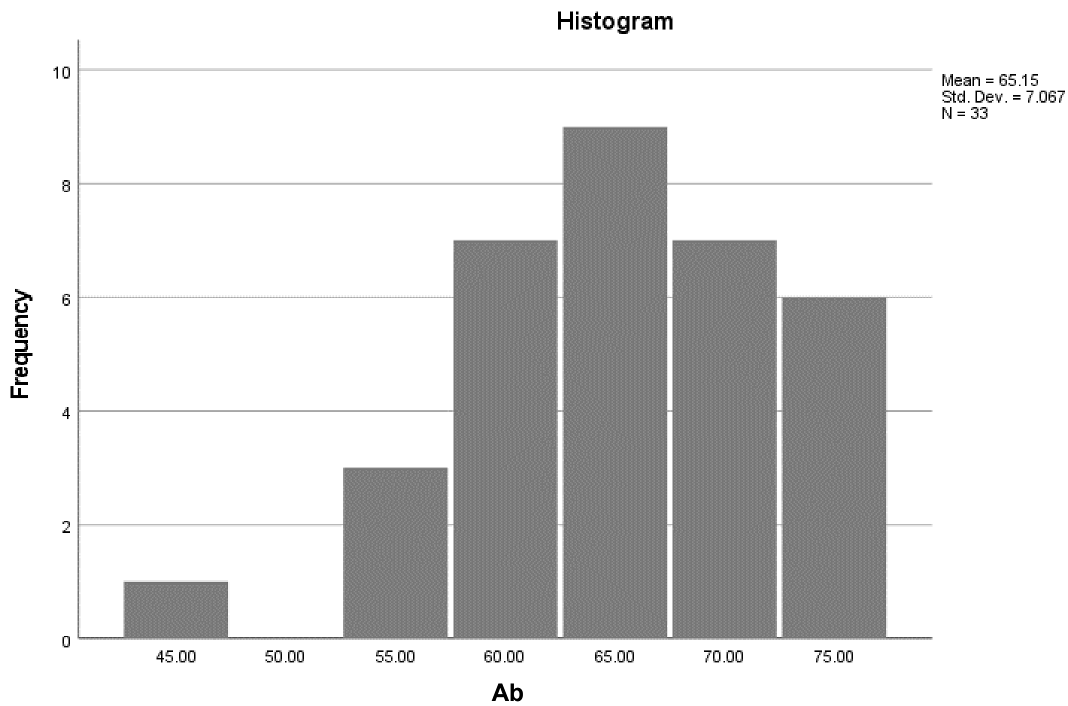


Figure S9

Boxplot of Participant Performance in Ineffective or Unallowed Questions Measure

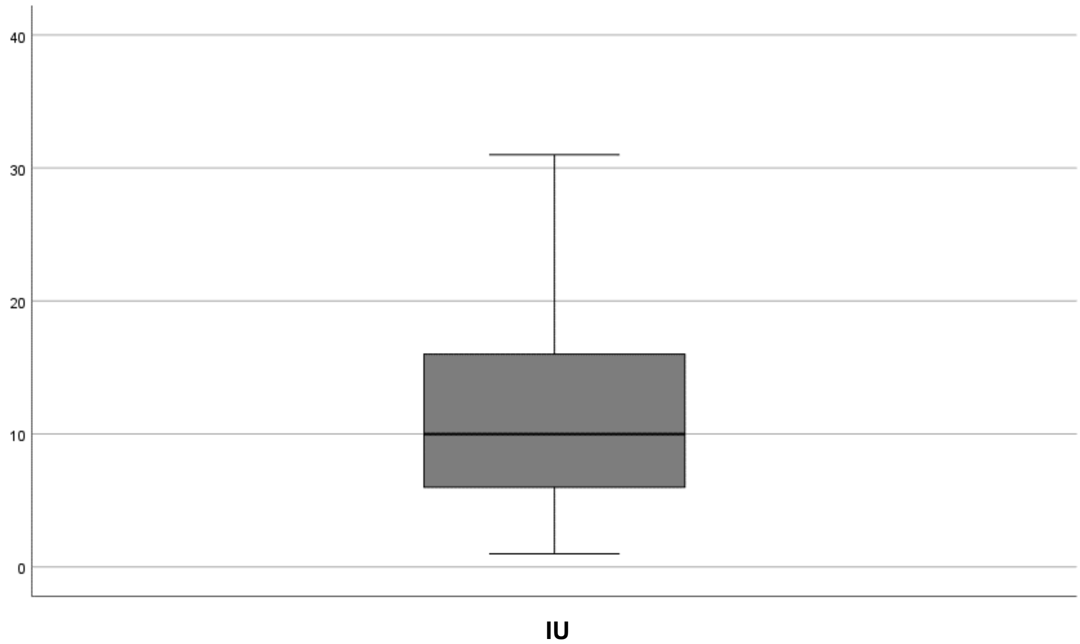


Figure S10

Histogram of Participant Performance in Ineffective or Unallowed Questions Me

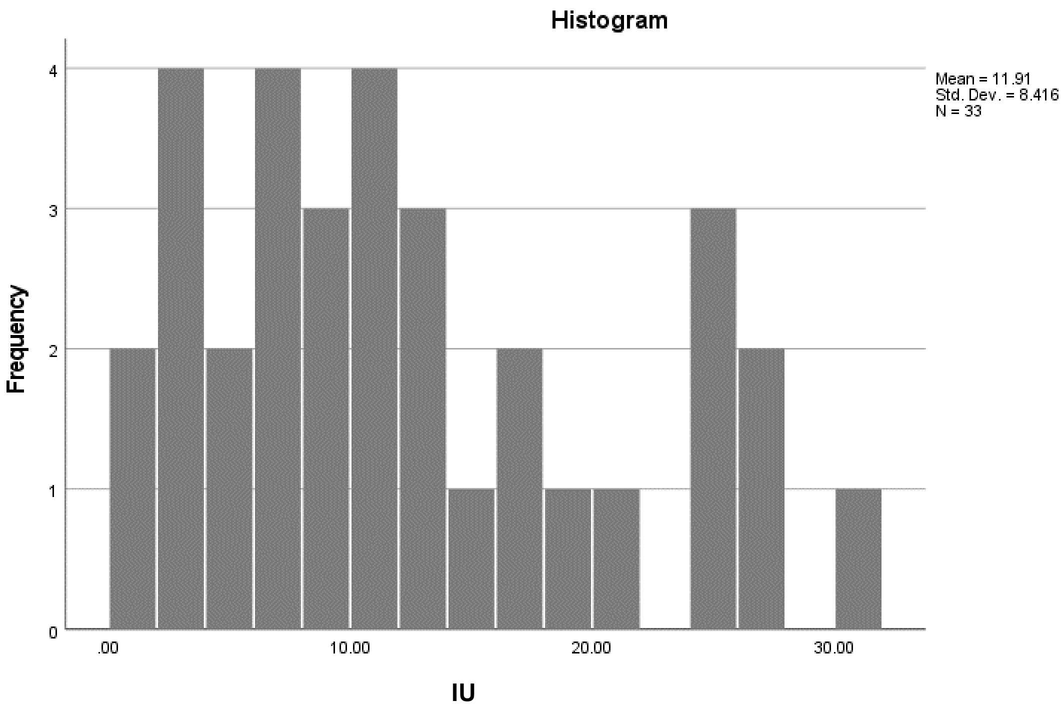


Figure S11

Boxplot of Participant Performance in Weighted Ineffective or Unallowed Questions Measure

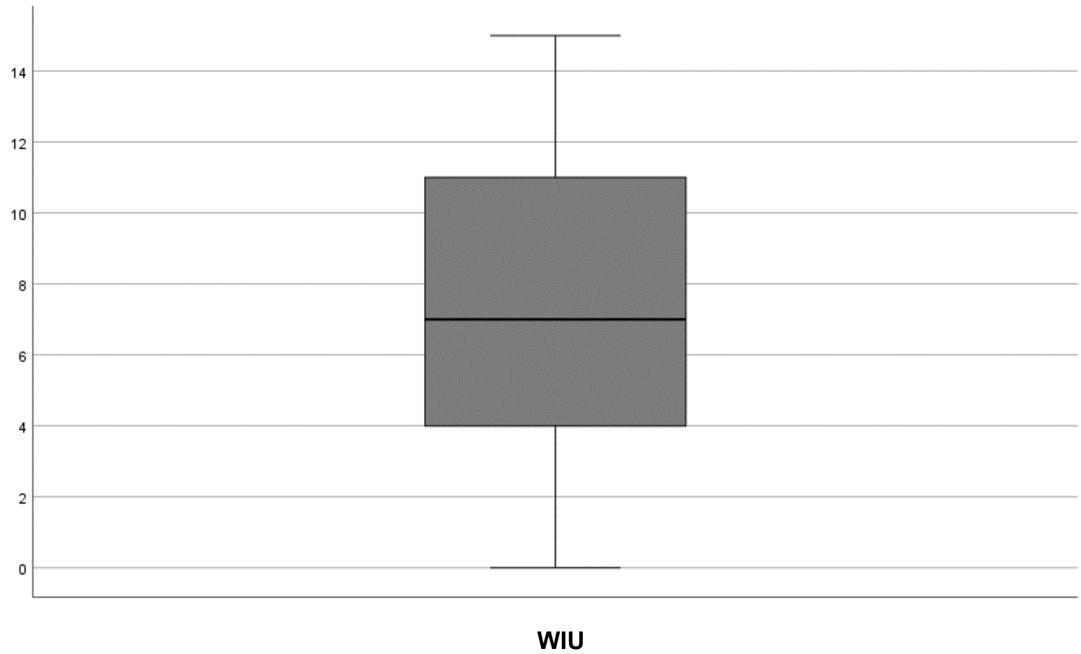


Figure S12

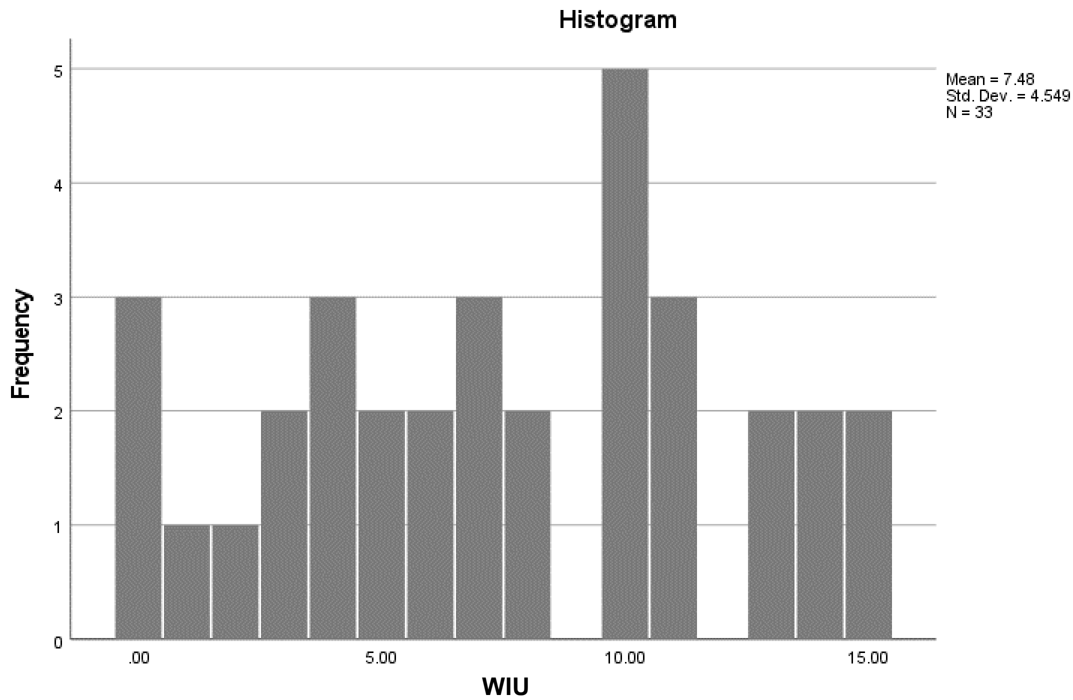


Figure S13

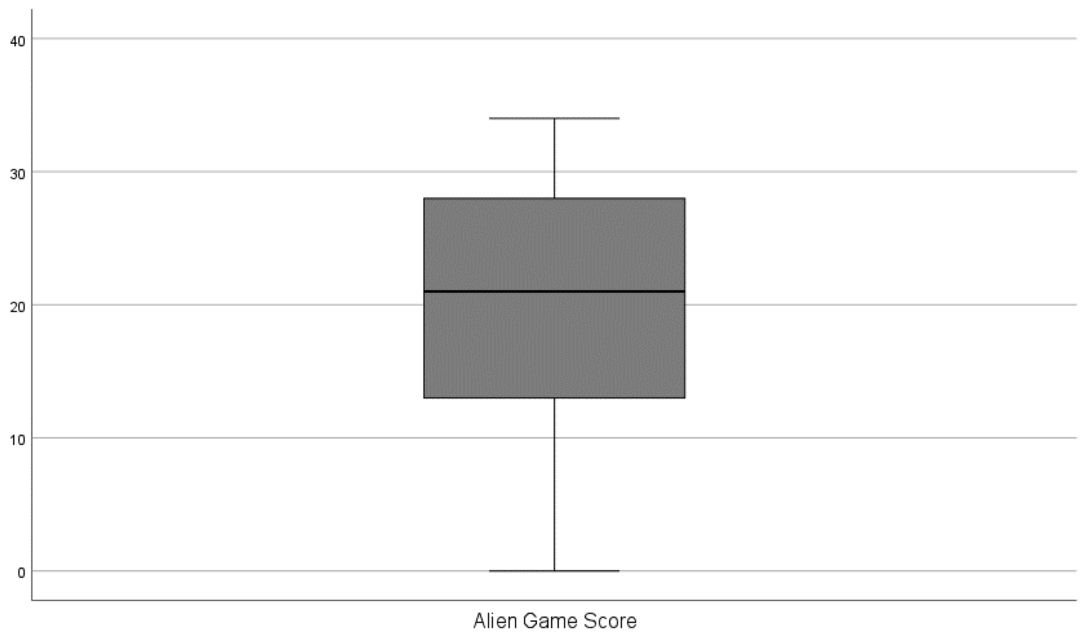


Figure S14