Using a Game-Like Task as an Assessment of Concept Formation in Children

Pinar Marasli

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

May 2023

ABSTRACT

This study aimed to develop Pavitt's (2017) newly developed game-based measure of concept formation and provide a practical scoring system for the Alien Game, with neurotypical children aged 8 to 11 years. Cross-sectional correlational design was used to compare the performance of participants on the Alien Game to established measures of concept formation (i.e., WISC-IV Similarities and Matrix Reasoning) and objective measure of executive function in everyday behaviour (i.e., Childhood Executive Function Inventory). Spearman's rank correlations indicated that the Alien Game was found to be a valid measure of non-verbal abstract reasoning. There were no associations found between the objective measure of adaptive function. There were also no group differences found based on sex and language. The Game was rated highly enjoyable by participants, and suggestions were made regarding how the game could be developed further. Results suggest that the updated Alien Game has the potential to be a suitable measure of concept formation for young children. Future research could develop the Alien Game based on the ideas given by the participants, and aim to recruit a more diverse sample of children with varying abilities.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my supervisor, Dr. Matthew Jones-Chesters. Without your continued support and guidance, this thesis would not have been possible. I am thankful for your kindness, reassurance and containment offered over the past three years. It certainly felt like I was in safe hands!

I would also like to extend my gratitude to my co-researchers for making this journey more bearable. A special thanks goes to Emily Hay for keeping everything under control; it was invaluable to have your support throughout this process; thank you!

I would also like to acknowledge the participants who took part in this research. Their willingness to contribute to my research was much appreciated. I would also like to thank the SENco of the school for being so welcoming and doing all that she could to ensure a smooth recruitment process.

My final thanks are extended to friends and family. A special thanks go to my husband for his continued support over the years, and for always helping me remain positive.

TABLE OF CONTENTS

List of Tables and Figures

LIST OF ABBREVIATIONS

ADHD	Attention-Deficit Hyperactivity Disorder
AG	Alien Game Score
Ab	Abstraction score
ASC	Autism Spectrum Condition
CHEXI	Childhood Executive Function Inventory
D-KEFS	Delis-Kaplan Executive Function System
EAL	English as Additional Language
EDA	Exploratory Data Analysis
EPL	English as Primary Language
IA	Initial Abstraction Score
TQ	Total questions
HS	Hypothesis-seeking Questions
IQ	Ineffective Questions
SES	Socio-economic Status
WA	Weighted Achievement Score for Total Questions
WCST	Wisconsin Card Sorting Test
WI	Weighted Score for Ineffective Questions
WM	Working Memory
WISC-IV	Wechsler Intelligence Scale for Children Fourth Edition

1. INTRODUCTION

Chapter one is an overview of the executive functions and concept formation and the associated deficits. It then provides a brief summary of current tests of concept formation and their limitations, and how a game-based approach could address some of these problems. A review of literature on assessment methods that have utilised a gamified approach and the current aims and questions of this research is also presented. Chapter two includes the methods and procedure used in data collection, and chapter three outlines the findings. Chapter four provides a summary of the findings and discusses the conclusions, implications, and suggestions for further research.

1.1. Executive Function

Executive function is an umbrella term used for a set of cognitive processes that are needed for adaptive and future-oriented behaviour. It is believed that the executive function is involved in many of the "higher-order" cognitive processes that subtend a range of skills such as paying attention and remaining on task (Micalizzi et al., 2019); problem solving; and considering dilemmas from various perspectives; knowing how one thing relates to another; and exerting self-control (Zelazo et al., 2016). These cognitive processes allow us to achieve a goal through our capacity to "plan, organise, and monitor the execution of behaviours that are strategically directed in goal-directed behaviour" (Micalizzi et al., 2019, p.3). Therefore, executive functions are utilised when it would be unwise to go on 'autopilot' precedence or depend on intuition; for instance when faced with novel or complex situations (Diamond, 2020a).

According to the influential conceptualisation of Miyake et al. (2000) there are three fundamental mechanisms of executive functions, which include working memory, inhibitory control, and cognitive flexibility – all three consist of a verbal and a visual subpart. Working memory helps us to actively hold information in mind, that we have seen or heard, and make sense of the information and decide how one part relates

to others (Diamond, 2013). Inhibitory control helps us to suppress or delay a dominant response that would impair or delay the attainment of a specific goal (Wiebe et al., 2011). Cognitive flexibility, which is also sometimes referred to as set-shifting or mental flexibility (Diamond, 2013), plays a critical role in our ability to fluently switch between several tasks in response to everchanging environmental stimuli (Wiebe et al., 2011). The literature in this area suggests that the complex cognitive and sub-processes involved in executive functions are mediated by the frontal and prefrontal cortex (Morris et al., 1993) and connections.

Executive functions are regarded to play a significant role in children's emerging emotional, social, academic, and behavioural competencies (Howard et al., (2015), and are considered to be a better predictor of school achievement than traditional IQ (Blair & Raver, 2015). For instance, the literature highlights that children's executive functions correlate with school preparedness (J. A. Welsh et al., 2010), literacy, and numeracy achievements (Clark et al., 2010); and it also predicts socioeconomic status and the likelihood of judicial involvement in adulthood (Huffman et al., 2001). Given that executive function skills are important for all aspects of life, early identification of difficulties in this domain of cognition, and appropriate interventions, can help improve academic and social outcomes (Bierman et al., 2008).

1.1.1. Deficits In Executive Function

It is widely recognised that executive functions are impaired in people who sustain damage to the frontal lobes; and those with neurodevelopmental conditions which are a result of congenital deficits affecting the frontal lobes (Hill, 2004). The diverse difficulties associated with deficits of executive function include problems with self-regulation (Sokol & Müller, 2007), social competence (Hughes & Ensor, 2007) and attentional difficulties (Martínez et al., 2016). This is also apparent in developmental disorders such as autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD); (Willcutt et al., 2005). A review conducted by Hill (2004) highlights that people with ASD, when compared with typically developing controls, exhibit impairments in flexibility and planning, and have difficulties with response inhibition similar to the difficulties experienced by those with ADHD (Sergeant et al., 2002). Further, people with intellectual disabilities also show features that are indicative of

executive function deficits. These include impaired self-control, mental rigidity, inability to implement activities, and poor self-hygiene behaviours (Griffith et al., 1999).

As noted above, executive function deficits can also occur due to a number of neurological conditions resulting from acquired brain injury (e.g., heady injury, bacterial meningitis, stroke), which may contribute to severe and longstanding disabilities in instrumental daily living activities (Chevignard et al., 2010). Acquired brain injury is one of the main causes of disability in children across the world. Although the majority of such children are likely to make good recovery and only around five percent may be severely impacted by their injury, an acquired brain injury following a traumatic brain injury (e.g., resulting from a road traffic accident) could have long lasting effects on a child's psychological, cognitive and educational functioning (Palanivel & Burrough, 2021). Consequently, cognitive assessment testing can play a significant role in the identification, rehabilitation, and promotion of recovery from injury (Gerring & Wade, 2012).

The executive functions can be usefully separated into the receptive (i.e., concept formation, also referred to as abstraction or induction) and expressive (i.e., task setting and task switching) aspects. The present study is concerned with the receptive aspect of executive function.

1.2. Concept Formation

Concept formation is a complex cognitive process that involves a range of cognitive functions such as abstraction, categorisation, and the ability to originate semantic-linguistic connections. It also provides a set of cognitive representations that underpin a range of complex thoughts, actions, and behaviours (Medin & Smith, 1984). Although there is no clear agreement on the general definition of concept formation, the most widely accepted understanding is that it consists in one's ability to categorise (put together) and differentiate stimuli based on their relational, perceptual, and functional features (Guthrie et al., 2004). Osgood (1953) stated that when humans think about concepts this can be done either in a concrete (e.g., tree)

or abstract (e.g., gravity) way. Ku (2019) explained concept formation as the ability to infer patterns about the world, based on existing knowledge, which allow generalisation from limited experience to novel exemplars (Ku, 2019). It has been postulated that concept formation plays a vital role in information processing and underpins various higher-order cognitive and language functions involving abstraction, reasoning, control of attention, and the ability to compare information (e.g., based on perceived properties such as physical size) (Vygotsky, 1986).

Kagan (1966) considered ‘concepts’ as being the prerequisite of human intelligence; and postulated that concepts are universal building blocks of cognition. Hence, conceptual development is regarded as being important for children’s development (Bracken & Cato, 1986). Indeed, fluid intelligence has been defined as our ability to solve problems, reason, and find relations and patterns among items (Ferrer et al., 2009), it makes sense why measures of fluid intelligence (e.g., the Raven’s Matrices; Raven, 2000) are often equated to executive function skills, such as concept formation.

Research also notes the role concept development plays in children’s language acquisition (Shaywitz & Shaywitz, 2005) and how this provides young children with the necessary vocabulary to make sense of conversations and instructions (Boehm, 2004). It has been well documented that concept skills are correlated with children’s school readiness, and academic performance (Panter & Bracken, 2009).

Literature in this area also provides evidence of the general importance of concept formation in supporting children’s overall development (Boscovich, 2006) and it is a better predictor of children’s mathematics and reading skills than the vocabulary tests usually used in clinical settings (Bull & Scerif, 2001). Given the crucial role concept acquisition plays in children’s overall development, it is important that measures designed to assess conceptual skills in children are suitable and sensitive, so that difficulties in this area can be accurately identified, and appropriate support can be offered (Wilson, 2004).

1.2.1 Concept Formation in Children

Although it is beyond the scope of this article to offer a comprehensive summary of the various theories of concept formation, Piaget's Theory of Cognitive Development (1965, 1964) in children will be briefly described because of its comprehensiveness and application to issues concerning education (Klausmeier & Hooper, 1974).

Piaget claimed that children's thinking is different from that of adults (Gallagher & Reid, 1981) and that their concepts pass through a series of four stages in cognitive development from infancy through adolescence: the sensorimotor, preoperational, concrete, and formal organisational stages. This theory focuses on how development takes place in sequence. Further, each of these stages are claimed to give rise to relating behaviours and operations. Piaget also postulated that children gain knowledge about the world and their sense of self through their interaction with their environment, and thus places an emphasis on environmental influences on cognitive development.

Children below seven years of age were administered Piaget's 'conservation' error' (Piaget & Szeminska, 1952) task, where liquid was transferred from a wide glass into a tall, narrow glass. Most children seemed to fail this task and were convinced that the taller glass contained more liquid. Given the importance of number conservation for a child's mathematical development, the conservation errors have been studied extensively. However, there are also shortcomings of the study. For example, Rose and Blank (1974) found that most six-year-olds successfully completed this task in their version of this study. It may be that children failed these tasks not because they lack the ability to grasp the number conservation principle per se but rather because the children may have inadvertently assumed that the adult intentionally changing the appearance of the liquid was important and so gave an answer which was impacted by experimenter demand (Pavitt, 2017). In a study conducted by McGarrigle and Donaldson (1974) they created a task where the modification was 'accidental' in their 'naughty teddy tasks'; where children were presented with two identical rows of sweets lined up, and a naughty teddy 'accidentally messing up' one line. It was found that, more than half of the children aged 4-6 years were able to correctly identify that the number of sweets had not changed, indicating that children are in fact able to conserve at an earlier stage than Piaget suggested.

Despite the shortcomings, Piaget's theoretical constructs could prove useful in considering the developmental norms of concept acquisition in children (Klausmeier & Hooper, 1974). This could enable adults (e.g., teachers, health professionals) involved with children's development to notice difficulties and support children to improve their conceptual skills and make the necessary adaptations to the teaching material to optimise learning (e.g., offering differentiated instructions in mathematics).

1.2.2. Role of Play and Games in Concept Formation

Professionals working with children recognise the importance of play and games in supporting children's developing minds. Play not only enables children to acquire the foundations of abstract thinking, self-reflection, and emotional regulation, but it also helps them acquire intricate communication and meta-communication skills and learn to discover and adjust to the roles and rules of functioning in society. For instance, engaging in some form of action and in an imaginary situation (e.g., pretending to drink tea from a shell as if it was a mug), and in the symbolic use of objects (e.g., using a skipping rope as a hosepipe), is a powerful vehicle for fostering children's emotional development and communication (Wieder, 2017). Hence, sociocultural theorists highlight the significance of play in child development and consider it to be one of the fundamental activities in individuals' formative years (Vygotsky, 1977; Bodrova & Leong, 1996).

Contemporary theories of play also focus on the emotional domain of child development. For example, Anna Freud (1968) and Erikson (1963) characterise play in terms of catharsis (i.e., allowing the expression of negative emotions, such as anger, in a safe context). Further, play provides children to experience novelty, uncertainty, and complexity. Such exposure enables children to integrate new and familiar experiences, which is also important for concept formation. As such, it is not surprising that play is located at the heart of under-five settings (e.g., play centres for toddlers).

1.2.3. Importance of Concept Formation in Children

Concept formation plays an important role in a child's academic skills as it is needed for various learning activities such as mathematics, reading comprehension and written expression. For instance, conceptual knowledge is important in reading because it enables a child to find intricate relationships between different ideas and storylines (Guthrie et al., 2004; Langer, 1967). With regards to writing, conceptual skills help children not only logically organise their ideas but also assist with the writing process. The ability to engage in mathematical problem-solving tasks also relies on the child's understanding of different concepts (e.g., subtraction and multiplication). Research in this area highlights that children's early literacy and mathematical conceptual skills are strongly predictive of their long-term academic success (Duncan et al., 2007), employability and income in adulthood (Rivera-Batiz, 1992).

Booth and Waxman (2002) found that children with strong conceptual skills were better at generating meaning when they were presented with novel words. Similarly, Mintz (2005) found that children as young as three years old could classify novel adjectives into appropriate categories when they were able to draw from their prior knowledge. It is also documented that skills associated with executive functions help with day-to-day activities required in social, academic, and professional settings. For example, the ability to create an exam revision schedule and transfer ideas onto a presentation depends on executive functions (Gioia et al., 2000).

Concept formation also plays an important role in developing social skills. For instance, to interact in a socially desirable way, the person needs to understand and relate to other people's viewpoints and have the necessary communication skills to share one's own ideas. Therefore, concept formation is closely linked to behaviours needed in everyday life and are vital in our cognitive function. Accordingly, most intelligence test sets incorporate a measure of concept formation (Hammill et al., 2009; Wechsler et al., 2004) in their assessment.

Given that concept formation is involved in everyday adaptive functioning skills (Gligorović & Buha, 2013) and are utilised across a wide range of academic, cognitive and social domains, it is important to assess concept formation in both

clinical and non-clinical populations. It is particularly important to assess concept formation skills from a young age so that the relevant support and interventions can be offered to children in a timely manner. Despite this, there are few widely available tests that assess concept formation in young children (Gelman, 2009).

1.2.4 Deficits Associated with Concept Formation in Clinical Populations

There is an array of research that focuses on the diverse range of deficits associated with poor concept formation in people with a range of conditions. As it is not possible to offer a comprehensive summary of the existing literature, a few key areas will be outlined below.

1.2.4.1 Deaf children. Given that concept formation is mainly acquired through the person's interaction with their environment and sensory experiences (Gelman, 2009), as well as experiences that are facilitated by language (Nelson, 1996), children with language delays as well as sensory impairments may be at a higher risk of experiencing delays in their development of concept formation. Deaf children experience problems with recognising relational concepts, such as 'opposites' (Zevenbergen et al., 2001). For example, in a cross-sectional study conducted by Pettifor (1968), it was noted that deaf children were slower to achieve in Piagetian conservation tasks (Furth,1964). They also showed deficits in nonverbal visual-spatial tasks (e.g., the "Tower of Hanoi"); (Luckner & McNeill, 1994) and verbal problem-solving tasks (e.g., in the Delis-Kaplan Executive Function System 20-Questions; D-KEFS; Delis et al., 2001); (Marschark & Everhart, 1999). Castellanos et al. (2015) examined whether deaf children who used cochlear implants were different in their conceptual skills compared to children with typical hearing. Children were administered tests of concept formation, including the Matrix Reasoning subtest found in the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler,1999), and the concept formation subtest of the Woodcock-Johnson Tests of Cognitive Ability (Woodcock, McGrew, & Mather, 2001). It was found that although children with cochlear implants could recognise visual concepts based on a single perceptual characteristic (e.g., shape or colour), they found it more challenging to engage in 'higher-order' cognitive processes involving relational concepts (e.g., combining or integrating similarities and differences). There were also deficits in their

verbal working memory and the speed at which they processed the information. However, this population of children may perform better in measures of concept formation that are nonverbal in nature (Castellanos et al., 2015).

1.2.4.2. Children with autism. It is noted that people with ASD often have an intense preoccupation with limited categories of information (e.g., with bus routes, dishwashers, or toothbrushes); (Firth, 2003). These limited interests are grouped under the restricted interests and repetitive behaviours triad of impairments reported in ASD (Wing & Gould, 1979). These behaviours also appear to be reflective of other related characteristic of ASD, such as giving close attention to detail and being highly receptive to small changes in the environment (Kanner, 1994). It has been suggested that organising new information is dependent on our ability to cope with differences and similarities, and forming groups based on perceptual or conceptual associations. As such, categorisation, is regarded to be pivotal to information processing (Rosch, 1978). Kingler and Drawson, 1995) note that difficulties in deviating from similarity or fixations on specific categories, could help explain the difficulties people with ASD have in categorising information.

Several studies have looked at how people with ASD learn and recognise categories, and how this translate in their goal-directed behaviours, and problem solving. Some studies have identified differences in the way people with ASD attend to perceptual similarities (Plaisted et al., 1998) and in how they generate concept prototypes (Klinger & Dawson, 2001). However, other studies have shown that the ability to derive new categories were intact in people with ASD when they were presented with basic shapes of categories (Bott et al., 2006; Soulieres et al., 2007), and were able to do as well as typically-developing individuals in their free sequencing (McGonigle Chalmers & Alderson-Day, 2010) and basic sorting abilities (Tager-Flusberg, 1985; Ungerer & Sigman, 1987). Nevertheless, research highlighted that when presented with more complex or abstract concepts, there was a significant drop in their ability to sort items based on conceptual categories (e.g., flowers, tables, or dogs) and perceptual characteristics (i.e., shapes and colours). Further, people with ASD also appear to show preference for concrete sorting (e.g., hat type) even if the correct answers were based on an abstract category (e.g., facial expressions); (Ropar & Peebles, 2007). This was also documented in a number of

other studies (Weeks & Hobson, 1987; Shu et al., 2001), where people with ASD completed fewer categories on the Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993) when sorting cards based on abstract relevance.

Parallel difficulties were observed on the D-KEFS 20-Questions Test (Delis et al., 2001); a measure of concept formation which requires the person to form abstract categories and change their questions based on feedback while they hold in mind their previous question. The aim of the task is for the participants to identify the target item in a maximum of 20 questions. Compared to typically developing children (TDC), children with ASD completed fewer trials (i.e., they were unable to identify the target item in 20 questions) permitted and asked fewer 'constraint-seeking' questions (i.e., questions targeting a category, or a feature shared by multiple items) and asked more 'hypothesis-seeking' questions (i.e., questions that eliminate single items); (Minshew et al., 1994). This highlights that although individuals with ASD could engage in the process of learning and recognising simple forms of categorisation, they find it challenging to generate effective strategies to identify more abstract categories. An advantage of using the 20-Questions and similar tasks is that it enables the participants to engage in the task as they would do in a real-life situation. However, difficulties with the 20-Questions may be as a result of difficulties with other executive functions, such as poor working memory (Bennetto et al., 1996; Williams et al., 2006) cognitive flexibility (Hill, 2004), as well as difficulties with individual's language skills to communicate effective questions.

1.2.4.3. Patients with brain injuries. Patients with frontal lobe lesions are reported to experience deficits associated with weak concept formation and problem solving abilities (Goldstein & Levin, 1991; Stuss & Alexander, 2000). In a study by Baldo et al. (2004), patients with frontal lobe dysfunction were administered the D-KEFS 20-Questions Test (Delis et al., 2001). Findings showed that patients with frontal lobe dysfunction asked fewer abstract questions than controls, and more questions that failed to remove any items, illustrating that they adopted an ineffective elimination strategy. The patients focused more on single items (e.g., 'is it the knife?') and struggled to see the superordinate categories (e.g., vehicles) in the task, which would have allowed them to narrow down their search. So, they were more concrete in their reasoning, which resulted in them asking more questions overall. Some of

the patients also utilised “pseudo-abstract” questions (i.e., a type of question that give the impression of a higher-level question, but only focuses on a single item. Despite this, these findings were consistent with other studies noted above.

The following section will first outline the current tests of concept formation before referring to the wider problems associated with current tests, and how the idea of gamification could help overcome some of these problems.

1.5. Current Tests of Concept Formation

This section is concerned with a brief overview of the current methods used in the assessment of concept formation in children. Summarising the assessment methods used in this area will situate the present study in its broader context and provide an insight into the importance of developing assessment methods that are suitable for young children.

Pavitt (2017) noted that tests of concept formation come in either a single-trial or multi-trial format. Single-trial formats provide participants with all the information required to complete the task within an item trial (e.g., the examine is presented with a single question/task, and the examinee could attain the correct answer by simply reviewing the options presented to them), whereas multi-trial formats require participants to build their answers through multiple attempts and feedback (e.g., the examiner presents the examinee with a single question/task, and the examinee is more likely to attain the correct answer by asking several questions). Both the single-trial and multi-trial formats contain components that either focus on the verbal or visual abilities of the individual.

1.5.1. Single Trial Tests

1.5.1.1. Visual single trial tests

There are a number of single-trial tests that are visual in nature (e.g., examinees are provided an upper array of conceptual categories with response options below). For instance, the Cattell Culture-Fair Intelligence Test (Cattell & Cattell, 1973), which is used with children aged four through to adults. Raven’s Progressive Matrices

(Raven, 1936; Raven, Raven & Court, 1998) in the Leiter International Performance Scale (Leiter-3; Roid et al., 2013), the Matrix Reasoning subtest in the Wechsler Intelligence Scale for Children (WISC-V, Weschler, 2014) and in the Wechsler Adult Intelligence Scale (WAIS-IV, Weschler, 2008).

Such tests are often assumed to be culturally fair, and appropriate to use with non-English speaking people as they require people to look at abstract figures and see relations between them (Miller,1966). Because this is a non-verbal test in nature and the participants are not required to share their reasoning, it is often assumed to be culturally fair. However, researchers state that the task could still be confounded with the participants verbal mediation skills (Sattler & Dumont, 2008) and that visuo-spatial abilities could differ across cultures (Gonthier, 2022). For example, research demonstrated that people from African cultures preferred colour over shape when they were shown abstract images (Serpell, 1979); and children from White middle- and upper-class backgrounds scored higher on the Cattell Culture-Fair Intelligence Test when compared to children from working-class backgrounds (Miller, 1966).

1.5.1.2. Verbal single trial tests

Verbal single-trial tests of concept formation include the Similarities task used in the WISC-V and WAIS-IV batteries (Wechsler, 2008). In this task, participants are asked to say how two things are similar or different (e.g., how are sheep and cows alike?), to assess their verbal reasoning. A limitation of this task is that it relies on language abilities. Other tests of concept formation include the Proverbs task (Gorham, 1956) a version of which is provided in the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001); and the Analogies task, found in the Wide Range Intelligence Test (Glutting et al., 2000). Although the use of analogy and proverb tasks taps into abstract concepts (Urbanski et al., 2016), performance on these tasks are confounded by people's general knowledge and vocabulary.

1.5.2. Multi Trial Tests

Pavitt (2017) notes that the most widely used multi-trial tests of concept formation are the concept formation subtests found in the D-KEFS and Woodcock-Johnson batteries (Woodcock et al., 2001). Pavitt notes that multi-trial tests may be more

ecologically valid than single-trial tests as it enables the examinees to engage in problem solving skills as they would do in real life.

1.5.2.1. Visual multi-trial tests

Among visual multi-trial tasks, card-sorting appear to be one of the most widely used formats. These include the D-KEFS Sorting task (Delis et al., 2001), the Weigl Color-Form Sorting Test (Weigl, 1941), and the Wisconsin Card Sorting Test (WCST; Heaton et al.,1993). Each of these tasks come with a different set of instructions; it may, for instance, instruct the examinee to correctly sort cards based on their distinct categories (e.g., visual-spatial, or verbal-semantic characteristics) and explain the strategy they have adopted (Delis et al., 2004).

The WCST seems to be one of the popular tests for assessing concept formation in children (Huizinga & van der Molen, 2007). In the WCST, examinees are shown stimulus cards and instructed to match the cards. The WCST was noted to be good at capturing the switching and inhibition side of executive functions (Gilgorovic & Buha, 2013). Patients with frontal lobe impairments are reported to find this task more challenging and make more perseverative errors (Teubner-Rhodes et al., 2017). However, research has also questioned its construct validity and posits that it may not be sensitive enough to pick up difficulties related to frontal lobe dysfunction (Bowden et al., 1998).

The Brixton Spatial Anticipation Test (Burgess & Shallice, 1997) is a relatively new measure for testing the ability to abstract, follow, and switch rules; and has resemblances with the WCST. Although the WCST is suitable to use with both children and adults, the Brixton task is not considered to be an appropriate test to use with children as it has not been normed with children (Pavitt, 2017).

Another visual multi-trial test of concept formation is found in the WJ-III-Cog test. The concept formation task in this test requires frequent rule application and switching from one rule to another; and thus requires good problem solving skills. Although some of the subtest in the WJ-III-Cog has been reported to have low subtest reliability and low construct validity (Campbell et al., 2008), a strength of this

test is that it was standardised using a sample across a wide age range (2-to-95 years), making it suitable to use with children (Edwards & Oakland, 2006).

1.5.2.2. Verbal Multi-trial tests.

In these tests' examinees are required to derive concepts from the information presented by the examiner.

An example of a verbal multi-trial test of concept formation is the D-KEFS 20-Questions which can be used with people aged 8-89 years (Delis et al., 2001). A strength of the D-KEFS 20-Questions is that it had a reliable convergent validity with WJ-III-Cog tests when tested with a large sample of children and adults (Flyod et al., 2006). This test was originally devised from a game that invites individuals to guess the name of a famous person by asking as few yes/no" questions as possible to guess the correct answer (e.g., "Is the name Andy?"); (Baldo et al., 2004) and later adapted in neuropsychological tests to assess executive functions skill (e.g., planning and monitoring information while the examinee holds in mind the responses from the examiner); (Anderson, 1998).

Similar to Mosher and Hornsby's (who were to first people to adapt the 20-questions task to use with children) (1966) test, the D-KEFS 20-Questions presents examinees with a set of 30 pictured items and requires them to guess the target items in no more than 20 questions. The responses from examinees are classified into various codes to capture the strategy used by examinees. These codes include: 'constraint seeking'; 'pseudo-constraint seeking'; and 'hypothesis seeking'. This test aims to assess concept formation through the examinee's verbal responses and ascertain whether the examinee could adopt effective strategies to categorise the items into hierarchical structures (e.g., super- and subordinate categories). Examinees are asked to eliminate the items/categories by asking as few yes/no questions as possible, while they attempt to hold in mind their questions and the examiners answers, to generate their subsequent questions.

The D-KEFS appears to be a good battery to use with children as both the verbal and non-verbal tasks are norm referenced with children, and it has good test-retest reliability and internal consistency (Diamond, 2020). Moreover, the game-like nature

of this test may make the testing process more enjoyable for children (Baldo et al., 2004). It is, however, important to bear in mind that although this test has been normed referenced to use with children as young as 8 years, it is still a test that is more suitable for adults as it demands greater language skills; it is also a less sensitive measure of concept formation in pediatric medical populations (e.g., children with epilepsy); (Parrish et al., 2007).

In summary, visual tests appear to be more culturally fair than verbal tests, as the latter is more influenced by the person's language skills and cultural background. Although there seem to be good measures of executive functions on the expressive side (i.e., the Wisconsin Card Sorting Task, which requires switching), current tests of concept formation and hypothesis testing are not very appropriate to use with children. For this reason, this present study is concerned with the development of a test that could potentially be used in the future as a multi-trial test of concept formation in young children.

1.6. Problems with Current Tests

The general limitation of current tests of concept formation is that it is unclear if they are suitable for children's developmental age, as the methodological approaches utilised to develop these tests have primarily focused on adult populations (Cirino et al., 2018). In other words, the formats and norms have been derived from adult populations and do not have extensive child norms (Welsh & Pennington, 1988). Further, most tests lack ecological validity. Given that the main goal of psychological science is to make sense of human cognition and behaviour and overcome the 'real-world' versus 'lab-dilemma' (Holleman et al., 2020), it is important that measures used are ecologically valid and resemble real-life situations. It appears that multi-trial tests have a better ecological validity compared to single-trial tests, as they permit the person to solve the problem through multiple questions, which can help clinicians understand how the person may approach more complex situations in everyday life (Holleman et al., 2020). The ecological validity of a test is crucial when interpreting the findings and making clinical decisions about a person's care, particularly when

thinking about the implications it has for the person's well-being, education, social relationships, and employability (Tröster, 2000).

Additionally, the main approach to testing 'intelligence' thus far has been to test all individuals, regardless of their cultural background, with tests developed in Western countries. As such, a dominant criticism in the realm of neuropsychological tests is that they are designed for people from Western and middle-class backgrounds and thus the items used in tests unsurprisingly tap into the goals, motivations and values held by people who belong to this group. This is supported by research which show that the socioeconomic status of children – which is most often based on parental occupation and family income (Noble & Giebler, 2020) –predicts performance on a number of neurocognitive tests, including tasks that assess executive functions (Last et al., 2018). Although working-class families worked as hard as middle-class families to provide their children with the same type of learning opportunities, it was middle-class families who possessed these resources (Wilson & Worsley, 2021). Group differences such as race, socioeconomic and education factors could therefore obscure true test performance. In light of this, when devising assessment tools, researchers need to ensure that they develop tests that are culturally fair and do not disadvantage children from non-Western and lower socioeconomic backgrounds (Miller, 1966).

Most current neuropsychological tests typically use paper-and-pen formats, which can be experienced as burdensome, frustrating and overwhelming by children (Lumsden et al., 2016). For instance, it has been noted that children with ADHD tend to underperform in executive function tasks, due to their difficulties with maintaining the level of attention and motivation needed to perform optimally on such tests (Sergeant et al., 1999). In other words, their performance seem to reduce the longer they spend on a task, and as executive demands become more effortful (Dekkers et al., 2017). Indeed, research indicates that children's attention levels, whether or not they have a neurodevelopmental condition or an acquired brain injury, are predictive of their ability to persevere in tasks that challenge them (Lunkenheimer et al., 2019). Given that executive functions are still maturing in young children, it is important that tests designed for children are not too taxing. Accordingly, creating shorter tasks

would help minimise boredom and disengagement related to difficulties with attention, and allow clinicians to get the best out of children.

Given that concept formation plays a crucial role in a child's developmental trajectory, it is vital to use good tests of concept formation with young children. Current tests of concept formation mostly utilise performance-based tasks where the individual is required to exercise their skills in the presence of an examiner in clinical settings (Harvey, 2012); and this has been the main way of assessing such skills for a long time (Zelazo et al., 2016). Further, it is clear that less biased and more culturally fair assessment methods are needed to better capture children's deficits and strengths. One way tests can be made less biased, more engaging, and culturally fair is by using 'game-like' formats. As such, the next section will outline the benefits of gamification in assessments used with children.

1.7. Benefits of Using a Gamified Approach

There is a large and growing literature on the use of gamified assessment methods with both clinical and non-clinical populations. This section will briefly explain the ways a gamified approach could address some of the problems with current tests of concept formation.

Games occupy a unique place in our lives; it gives us a sense of agency in a more fluid way than one might presume. That is, in playing a game, we immerse ourselves temporarily in an alternate agency. Play is also the direct opposite of seriousness, and as a result, it does not concern itself with truth or falsity. Huizinga (1955) posits that play encourages momentarily stepping outside of "real" life into an alternate sphere of activity: a "magic circle". In this magic circle, the individual does not encounter the consequences outside the magic circle; and, as such, we are more willing to take up new roles and motivations.

An advantage of the use of the word "game" is that it is a familiar term to children, and gamification can therefore be an effective strategy when administering neuropsychological assessments to children. Research shows that children with

ADHD are likely to be more cooperative and attentive in situations where a gamified approach has been used. This could positively impact the reliability and validity of test results (Cheng et al., 2019), especially considering that loss of interest or fatigue could result in inaccurately low scores. Given that neuropsychology is often concerned with cognitive processes and ways people can be supported once difficulties have been established, it is, therefore, important for tests to have face validity (Nevo, 1985) and for children to find these tasks engaging (Gioia, 2015). The motivational power of 'game-like' elements, could therefore improve user experience (Lumsden et al., 2016), and encourage persistence and achievement in tasks (Ritchter et al., 2015), by making participation in tests more implicit (Cerrato & Ponticorvo, 2017).

Gamified assessments may also be easier to administer by professionals in non-clinical settings (e.g., by teachers) as these methods may not require extensive training (Gómez-Tello et al., 2022) and be shorter in duration (Cerrato & Ponticorvo, 2017). These will help reduce the cost of production and increase accessibility to resources and support, especially in areas where services are scarce or involve a lengthy wait time (Bauer et al., 2012). Administering a game-like assessment in a child's natural setting (e.g., school), may help the child feel more relaxed and enable them to approach the task in a way that is more reflective of their 'real-world' behaviour (Song et al., 2020).

It is widely documented that children who speak English as an additional language, and children with language impairments, tend to produce low scores on tests, due to difficulties with understanding the instructions and articulating their ideas (Barkan & Bernal, 1991). As a result, children with language difficulties are disadvantaged by these traditional assessment methods. This is concerning considering that there are almost one million children in primary schools in England who speak English as an additional language (School Census, 2021). Even if children are primary speakers of English, they may still struggle with the standardised tests, particularly if they live in poverty, as these tests are regarded to be biased in favour of children from White, middle-to-upper class backgrounds (Clark, 1997). It is possible that using a gamified approach could help eliminate the linguistic demands of tests and create a more "level playing field" for children from a range of backgrounds (Lewis, 2001).

Gamification can also be used in assessments aimed at developing children's social and emotional skills (Nicolaidou et al., 2022). Social and emotional skills may include interpersonal and intrapersonal skills which encourage the child to understand and manage their emotions, show empathy for others, establish positive relationships, and engage in sensible decisions (Wu et al., 2020). As children grow older, their learning performance tends to improve through their increased ability to regulate their emotions and behaviours (Arguedas et al., 2006). Considering that aggression, irritability, and anger are among the most common reasons for child and adolescent mental health referrals (Sukhodolsky et al., 2016), games and apps can be a useful way to engage with children more generally. Additionally, gamification can also be used to foster children's educational creativity, such as musical creativity, which can help children with intellectual disabilities to reach the same learning targets as their typically developing peers (Wong, 2020).

Another important consideration is the anxiety associated with traditional assessment methods. For instance, the child may experience worry related outcomes (e.g., "what if I fail the test") and responses (e.g., experience tension and increased palpitations), resulting in below optimum scores (Cassady & Johnson, 2002). Indeed, research has found a moderate negative correlation between test anxiety and performance on neuropsychology tasks (McCarthy & Goffin, 2005). This has serious consequences given that anxiety levels could hinder performance and disadvantage children who may experience higher levels of anxiety under test conditions. Interestingly, game-based assessments seem to generate less anxiety (Kocadere & Çağlar, 2015), compared to traditional assessment methods (Turan & Meral, 2018). Perhaps this is because a gamified assessment is perceived as being more engaging, immersive, and less onerous than traditional assessments (Levy et al., 2018). Clinicians have a responsibility to create a non-threatening atmosphere when assessing children; a gamified approach could lessen performance anxiety.

As such, the present study hopes that using a game-like format will make the assessment of concept formation more engaging and interesting to children than the traditional "paper and pencil" tests used in clinical settings. The next section provides

existing research that has used a gamified approach in the assessment of concept formation in young children.

1.8. Literature Review Process

1.8.1 Method

A literature review was conducted to discover papers relating to the assessment of concept formation in young children using a gamified approach. A number of databases were explored including PsychINFO, CINAHL Plus, Academic Ultimate, Scopus, and PubMed. Google Scholar was also utilised to support the literature search process. The following search terms were used to identify suitable papers: 'Concept Formation' OR Abstraction OR Induction, along with Child* (where * indicates shortened terms) AND Test OR Assess* OR Measur* OR Exam*, AND Game* OR Game-like OR Game engagement OR Tablet-based OR Computer OR PC based formats.

1.8.2. Inclusion And Exclusion Criteria

1.8.2.1. Inclusion criteria

Studies were included if they met all of the following criteria:

a) Concerned with the receptive side of executive function (i.e., concept formation, abstraction, and induction)
b) Focused on the assessment of concept formation
c) Assessed concept formation using a gamified approach
d) Included primary school-aged children (even if the study included older children)
e) Were published in the English language
f) Published between 2000-2022

1.8.2.2. Exclusion criteria

Studies were excluded if they may any of the following criteria:

a) Examined the development or delivery of interventions for concept formation (not the assessment of concept formation specifically)
b) Examined concept formation skills solely using traditional assessment methods (not gamified methods specifically)
c) Examined concept formation in adult populations.

1.8.3. Search Results

The PRISMA Flow Diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) presented in Figure 1 notes the papers identified, screened, and evaluated for the literature review. The initial search identified 1,377 papers. After removing duplicates, the number of papers remaining was 1,371. Following this initial stage, the titles and abstracts were examined to identify suitable papers that met the inclusion criteria. References section of the papers that met the inclusion criteria were also examined to identify suitable papers. A summary of the findings of the literative review has also been presented in Table 1.

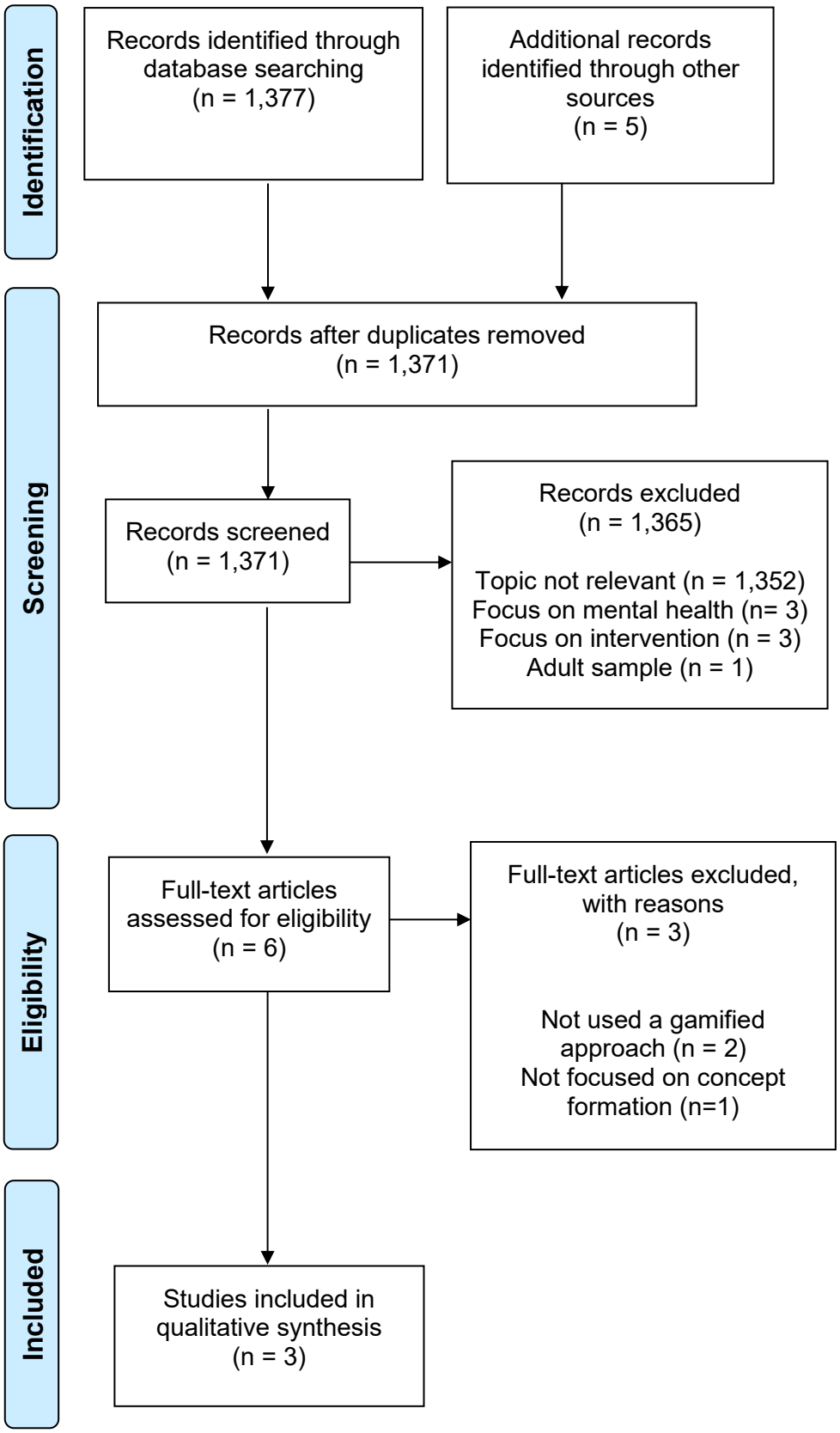


Figure 1. PRISMA (Moher et al., 2009) Flow Diagram of Article Selection Process.

Table 1

Details of Studies Included in The Literature Review (N = 3)

Study and location	Aim	Sample	Design	Gamified Approach	Summary of Results
Alderson-Day & McGonigle-Chalmers. (2011) Scotland	assess whether the children's concept formation skills, verbal IQ, and the test content impacted their ability to generate problem solving skills and form categories.	Total N = 28 Agre range = 8-17 N typically developing children = 14 N children with ASD = 14	Cross-sectional	Novel task consisting of 20 images of Robots (based on the D-KEFS Twenty Questions Task)	Typically developing children -asked greater abstract questions (i.e.., 'constraint seeking' questions -eliminated greater number of items Children with ASD -asked more questions when physical elimination was not possible on novel task -completed fewer trials -generally asked fewer abstract questions -struggled with working memory -verbal abilities predicted task performance
Condy et al., (2021) United States	establish the validity of the novel, tablet-based concept formation test, and whether the technology used had the appropriate interface,	Total N = 8 in the pre-piloting stage, and n = 15 in the piloting stage. Age range = aged 1 – 16 with a mental age less than 24 months	Cross-sectional	Tablet-based task	-control group and neurodiverse participants navigated the tablet-based test -Neurodiverse participants struggled more than the control group to mauver the tablet-based task
Pavitt. (2017) England	Whether the children could generate effective concept formation strategies in their attempts to problem solve, and if they could, what kind of strategies they utilised), and whether findings could help to establish 'response norms' in the future.	Total N = 18 typically developing children Age range = 7- 11	Cross-sectional	Alien Game (based on the D-KEFS Twenty Questions Task)	- Participants reached target Alien in 10 questions -Participants with higher Abstraction Scores (AS) asked greater 'constraint seeking questions' - Participants with lower AS scores asked greater ineffective questions (i.e., 'High risk' questions) -No significant learning observed through trials -Verbal abilities were more predictive of game performance

This section of the report outlines the qualitative synthesis from the 3 papers obtained in the literature search. It will describe the papers that have used a 'game-like' measure in the assessment of concept formation in children. It will also outline the strengths and weaknesses of all studies (e.g., sample size, ecological validity, ceiling effects). Finally, the reasoning and research aims for the current study will be offered. Since this current study aims to further develop Pavitt's (2017) study, a longer description of her study has been provided.

1.8.3.1. Is it a Bird? Is it a Plane? Category Use in Problem-Solving in Children with Autism Spectrum Disorders

Alderson-May and McGongle-Chalmer's (2011) administered two versions of the 20-Questions Task (based on Mosher & Hornsby's 1966 and D-KEFS 20 Questions Tasks) to a group of children with ASD (n=14) and typically developing children (n=14) ranging in age from 8-17. One task consisted of the standard version (which comprised of everyday items) and the other task consisted of a novel version (one which they developed consisting pictures of Robots) of the 20-Questions task. It was believed that the use of Robots would be more engaging and appealing to younger children. The aim was to assess whether the children's concept formation skills, verbal IQ, and the test content impacted their ability to generate problem solving skills and form categories.

The novel task consisted of a total of 24 images of Robots, with individual names, which could only be categorised based on perceptual features (e.g., shape, colour). Whereas the standard task could only be grouped based on conceptual categories (e.g., tools, animals, vehicles). Based on findings from previous studies, it was predicted that children with ASD would do better on the novel task more than the standard task, as they are more likely to base their search strategy on perceptual features. To reduce the load on the working memory, children were allowed to physically eliminate the items; similar the rule found in games such as '*Guess Who?'* (© Hasbro), where participants are allowed to knock over the eliminated characters displayed on hinged frames. To compare the effects of this on the working memory, they included trials that permitted and prohibited the use of physical elimination. Accordingly, they introduced three trial blocks to assess the flexibility of strategies

used by participants. These three trial blocks consisted of: an *initial* condition (whereby physical elimination of items by participants was permitted), a *flexibility* condition (physical elimination was permitted, but the distribution of categories varied), and a *memory* condition (physical elimination was not permitted). Finally, as the Robot task was highly verbal in nature, and the language demands could not be easily minimised, participants language skills were assessed using a verbal measure, to investigate whether their language skills had any role on their performance on the Robot task. Based on existing literature, it was predicted that the performance of children with ASD would be strongly influenced by both the executive and linguistic demands of the 20-Questions task.

It was found that participants with ASD completed fewer trials and asked fewer abstract questions than typically developing children on the standard task. Overall, children with ASD performed worse in the standard task, whereas on the novel task, there were no group differences (i.e., both groups asked a range of questions about the shape, colour, and other features of the Robots). However, it was not clear whether this was due to the 'ceiling effects' experienced in the novel task: the novel task was too easy for both groups of participants to draw any meaningful conclusions. Despite this, when comparing the quality and type of questions asked, children with ASD consistently asked more 'hypothesis seeking' questions than 'constraint seeking' questions, in both the standard and novel task, even when executive demands were reduced (for instance in the *initial* condition). This suggests that children with ASD utilised ineffective strategies of categorisation (i.e., they asked questions that removed fewer items) compared to their typically developing peers. Additionally, children with ASD also asked a greater number of questions than their typically developing peers when the elimination of items was not granted in the *memory* condition – suggesting that they also experienced challenges with working memory. Whereas typically developing children were better at asking abstract questions, and thus they adopted better problem solving strategies. Additionally, the performance of children with ASD was predicted by their verbal abilities, this was not observed for typically developing children.

A limitation of this study is that the participants were not matched on their verbal abilities, and although group differences were apparent in the question quality,

differences in vocabulary could account for the difficulties experienced in generating effective questions. Another important consideration is that the group differences could be due to the design of the tasks. While the standard 20-Questions task had items that made it possible to form groupings based on super-ordinate categories (i.e., animals, mammals), the novel task meant that groupings could only be based on the features included in the Robots (e.g., colour and shape). Since there were no group differences in the novel task, due to 'ceiling effects', future research could design more challenging tasks to minimise this from occurring.

Regardless of the limitations listed above, the study provides useful information about the way children with ASD approached categorisation on both tasks. It highlighted that the children with ASD adopted a more restrictive approach to categorisation, whereas typically developing children were less restrictive in their organisation of categories, which enabled them to ask more effective questions (i.e., 'constraint seeking' questions).

1.8.3.2. A Novel Measure of Matching Categories for Early Development: Item Creation and Pilot Feasibility Study

Condy et al., 2021 piloted a novel measure of concept formation, which was on a tablet (based on the NIMH Monkey-Logic 2 Toolbox®; Hwang et al., 2019), for children (aged 1 to 16) with a neurodevelopmental condition (i.e., those with a mental age less than 24-months). The aim was to establish the measures validity (i.e., whether the items tapped into concept formation), and whether the technology used had the appropriate interface, which required minimal verbal and motor input from children. Participants were asked to match items to semantic categories based on their logical relationships (e.g., things that you drink, things that are trees), and perceptual categories based on visual features (e.g., colour, size, and shape) of items using the drag and drop function on a tablet. The researchers first demonstrated the task to the children, before the children moved onto the practice items, and then onto the test items. The number of distracters increased in trials, and the task progressed from more basic to challenging categories; three different sets of varying levels of difficulty were included in the pilot-phase, to uphold participant interest.

The pre-piloting and piloting stages showed that the content used in the task could be developed further to create a test that taps into perceptual and semantic aspects of concept formation, albeit with some modifications (e.g., by adding more attributes and distractors). The task had acceptable to high levels of feasibility, respectively. The feasibility of the study was also demonstrated by participants ability to understand the instructions even when they were only given minimal verbal input (e.g., “See!”, “Your turn”) during the demonstration item, and in 80% of the children transitioning from the practice items onto the test items. Most, though not all, children showed good engagement (i.e., some managed to use the touchscreen in some way), and their motivation increased when they received feedback on their responses during the practice items, though their engagement decreased when feedback ended.

A strength of this study is that experts on the subject matter (i.e., psychologists, software developers) were recruited in the generation of test items and rules, as well as in the evaluation of the test interface. This helped create items with clear and appropriate task demands in line with the study aims. The pre-piloting stage helped researchers to gauge engagement levels, and whether the task had face validity by checking the “look and feel” qualities of the user interface, and whether the participants could physically manipulate the touchscreen.

The small sample size (n=8 in the pre-piloting stage, and n=15 in the piloting stage) is, however, a limitation of this study, which raises questions about the generalisability of the findings. With respect to the feasibility of using a tablet-based task by this population, even though the tablet-based task was developed to minimise motor demands. The most frequently reported challenge by the examiners was the difficulties children experienced in utilising the drag and drop function, causing them to feel frustrated. Given that young children can have attentional challenges, additional motoric demands may have further interfered with their engagement and performance. Some of the participants refused to use the device altogether, and others struggled to remember the instructions. Despite the limitations, this study provides useful preliminary information on the applicability of tablet-based tasks for children with a range of physical and/or neurodevelopmental difficulties (Hessl et al., 2016); Tulsky & Heinemann, 2017; Twomey et al., 2018).

The limitations of this study should be considered in future research aiming to develop similar technology-based assessment tools for young children with cognitive difficulties. For example, future researchers could minimise the motoric burden of a task and consider designing a novel task that does not rely on the children's ability to manoeuvre or operate upon a technological device.

1.8.3.3. Using a Game-like Procedure as a New Test of Problem Solving and Concept Formation in Children

Pavitt (2017) engaged in the first phase of an explorative approach aimed at developing a novel test of concept formation for typically developing children (n=18, ages 7-11). The focus of the study was to see if the children could generate effective concept formation strategies in their attempts to problem solve, and if they could, what kind of strategies they utilised), and whether findings could help to establish 'response norms' in the future.

Pavitt (2017) developed the Alien Game based on Mosher and Hornsby (1966) and Alderson-Day and McGonigle-Chalmers (2011), D-KEFS 20-Questions task, as well as on the propriety 'Guess Who?' game. To overcome problems associated with being familiar with the 'Guess Who?' game, the images were changed to cartoon drawings of 'Aliens' and assigned made up names. These changes helped minimise practice effects and were intended to make the test more culturally fair, as the use of Aliens were a novel stimulus to all children. Pavitt presented 24 pictures (as opposed to 30 pictures used in the standard 20-Questions task) of Aliens on a board with hinged frames, and the participants were considered to have 'passed' the trial if they guessed the target Alien within 10 questions.

In Addition to the Alien game, each child was also administered a standardised single-trial test of concept formation (i.e., the Similarities and Matrix Reasoning subtests in the WISC-IV battery). This was administered to identify verbal and visuo-spatial difficulties. Since the WISC-IV subtests utilise a single-trial format, and the Alien Game utilised a multi-trial format, the use of these measures did not contribute to practice effects. Similar to the 20-Questions task, the Alien Game was only played one way (i.e., the participants asking the questions and the examiner recording the

responses) during testing. Task enjoyment on all tasks were measured using a five-point Likert presented on a visual analogue scale.

Pavitt (2017) conducted a two-way content analysis to ascertain the quality of questions asked by participants. Following this, a final coding scheme was generated, and the inter-rater reliability was tested by a second rater, which helped to create a scoring sheet. Finally, an exploratory content analysis was run to explore the type and pattern of questions asked by the children. The minimum as opposed to the maximum number of items removed were used to calculate the quality of the questions. It was hoped that this would prevent the child from getting a high score because of a 'lucky guess'.

The categorisations were based on perceptual characteristics (e.g., shape and colour) rather than conceptual criteria. In addition to the codes used in the 20-Questions task (i.e., 'constraint seeking', 'hypothesis-seeking', and 'pseudo-constraint seeking' questions), Pavitt also introduced new codes to measure 'risk'. The 'risk' was coded as 'Medium risk' (i.e., a question that removed 6/24 items or less), 'High risk' (i.e., a question that removed 4/24 items or less). Pavitt also re-defined the traditional question codes in the following way: a 'constraint-seeking' code was given if a question eliminated two or more items; a 'hypothesis-seeking' code was given if there was a follow-up, taxonomical question, possible, but was not asked; and an 'Intradimensional' code was given if the child asked a taxonomical question based on their prior question. Finally, 'Failure to eliminate', 'Immediate repetition', and 'Delayed repetition' codes were also adopted. Failure to eliminate questions were questions unrelated to the previous question; immediate repetitions were repeated questions; and delayed repetition questions were repeated questions from an earlier trial.

To establish the type of questions asked by the children, an Abstraction Score (AS), a Learning Slope (LS), an Initial Abstraction Score (IAS), and a Weighted Achievement Score (WAch) were calculated. These scores were then correlated with the scores from the Similarities and Matrix Reasoning tasks. The scores on the Similarities and Matrix Reasoning helped ascertain whether the verbal and visuo-

spatial abilities of the children were correlated with their scores on the Alien Game. In addition, children's 'Time to Respond' to the first question was also measured. The first three questions asked in each trial, to calculate this score as some children completed a trial in three questions. The IAS and WAch were, however, scored in the same way as the 20-Questions scoring system. As such, the same strategy utilised by the 20-Questions was applied to determine the IAS score (this score was based on the fewest number of items removed). So, the only difference was the inclusion of an AS score in Pavitt's study.

In summary, the AS, IAS, and WAch scores were used to capture the normative performance characteristics – none of these scores were impacted by a potential 'lucky guess'. Results showed that a 'Constraint seeking' question type was the most effective strategy to use with the Alien Game, which also helped the children obtain a high score on the AS, IAS, and WAch. It was found that children with the highest AS scores only asked 'constraint seeking' questions, whereas children with the lowest AS scores would also ask questions coded as 'High risk' within that specific trial. However, there were no significant changes observed in the LS, and thus the children's performance remained consistent across the four trails in the Alien Game. The 'Time to Respond' score was not deemed to be a potential measure of normative performance characteristics as the results did not provide any consistent findings. Scores on the Similarities and Matrix Reasoning tasks indicated that verbal abilities were more predictive of scores, than visuo-spatial abilities. No correlation was found between the Alien Game and either of the WISC-IV subtests. This might have been as a result of the Alien Game being a multi-trial test in nature (and the WISC-IV subtests being a single-trial test).

All of the children managed to identify the correct Alien in a maximum of 10 questions, and they most commonly asked 'constraint seeking' questions. This highlighted that all children managed to hold in mind the rules of the game and adopted an effective strategy in their attempt to problem solve the task. 'Intradimentional' was the second most common question, meaning that children based follow-up questions on their previous question. It was also found that the children were unlikely to repeat a question in the trial, and if they did, they noticed

their mistake.

Although two questions (i.e., “Does it have no legs?”; “Does it have a tail?”) removed half of the items (i.e., 12 Aliens), this question was only asked once. Pavitt notes this might have been because these features failed to stand out to the children, compared to the other features that were more salient (e.g., the colour of the Alien). Additionally, although children were instructed to not ask a question about the names of Aliens, two children asked about the names.

A particular strength of this study is that the use of Aliens meant no child was more familiar with the images, and it potentially made the task more culturally fair. The findings also provide important information regarding the way typically developing children approached the test and used concept formation skills to aid their cognitive processes. Children adopted similar strategies in the Alien Game, which points to the potential for establishing possible normative characteristics in the future. Lastly, the ratings on task engagement showed that all of the children found the Alien Game more enjoyable than the Similarities tasks.

There were also some notable limitations of the study. For instance, the majority of the participants were from a White-British background and spoke English as a primary language, which limits the generalisability of the findings to other populations. In addition, participants scored in the above average range on the WISC-IV tasks, which is not representative of the general UK population. Further, the use of names on Aliens introduced a linguistic element to the task. Use of a boardgame format with hinged frames reduced working memory demands, and made the task more ‘game-like’, it also produced a ‘ceiling effect’ as the task was not challenging enough. Moreover, the game had a chance element as two questions removed half of the items (i.e., 12 Aliens), regardless of whether the question produced a correct answer or not.

Some of the recommendations outlined in this study include: improving the coding system to find out if there is a particular code that best captures children’s problem solving skills, but retaining the ‘Failure to eliminate’ code as this could help to identify children who may have deficits in their working memory; improving the images of the

Aliens; and comparing performance on the Alien Game against a more objective measure to capture children's executive function in the academic setting, for criterion validity. Pavitt also recommends researchers continue to build on standardising the Alien Game, so it could be used to assess difficulties in concept formation in younger and older children.

1.9. Summary of Literature Review

Overall, the literature shows that children with neurodevelopmental conditions engage in game-like assessment methods; however, further developments are necessary. Although children with ASD could engage in the process of categorisation, they generated less effective problem-solving strategies and asked fewer abstract questions, compared to typically developing children. This highlights the difficulties children with neurodevelopmental conditions have with concept formation, and the importance of identifying these difficulties at an early stage of development, so appropriate interventions can be utilised.

Pavitt's study indicated that children utilised similar types of questions (i.e., 'constraint seeking') and approached the Alien Game in a strategic way, making it possible to derive normative data and potentially standardising this as a formal test of concept formation for young children.

1.10. Present Study

1.10.1. Aims

The present study aims to develop the Alien Game further and administer this game to a new group of 8-to-11-year-old children. The use of Aliens has the advantage of making the task more novel and culturally fair than other tests. It was decided that targeting this specific age group would be more achievable given the time constraints of this study. It is hoped that the findings of the study would provide helpful information to future researchers aiming to target young children from other ages. Further, given that single-trial tests of concept formation lack ecological validity, creating a game-like test that is multi-trial in nature and one that can be

administered in the child's natural setting, would be more ecologically valid and enable the examiner to assess the child's real-world behaviour.

1.10.2. Rationale

Early identification and intervention of concept formation can help improve adaptive functioning, academic and psycho-social success in later life. However, current tests of concept formation are not suitable for young children (Cirino et al., 2018). Existing tests rely on Western culture constructs unsuitable for other populations. In summary, there is a general need for a measure of concept formation that is (a) engaging, (b) culturally fair, and (c) suitable for young children.

The current research aimed to continue the development of the Alien Game and contribute to the development of information and response coding system that will be used as a standardised test of concept formation in the future. The current study was tested on typically developing children aged between 8-11 years. Additionally, the current study aimed to develop the physical elements of the game, formalise the testing procedure by adding a game script and improve the scoring system.

1.10.3. A Game-Like Procedure

Firstly, a set of new 24 custom-drawn 'Aliens' (provided in Appendix A) were designed to make the game more appealing to younger children. The use of 'Aliens' were kept as this would be a novel stimulus for all children Features of the 'Aliens' were modified to make the characteristics more prominent, to avoid a single question from eliminating a large proportion of the items, whether the question yields a correct or an incorrect answer. The Aliens were designed to allow categorisation on perceptual characteristics (e.g., shapes and colours) given that this categorisation is less likely to be influenced by cultural factors. The names of the Aliens were also removed to reduce the linguistic element of the game. Instead, Alien's were given numbers, though this was only used to confirm the target Alien at the end of each trial. Instead of using a hinged board like in the 'Guess Who?' format, a low resource layout was utilised, to make the game cheaper to produce and more accessible for public use (e.g., for clinicians and teachers even in low resource communities and settings) in the future. Thus, the game was presented on A3 paper. This new design

may also minimise some of the problems associated with assessment methods (e.g., the use of a boardgame with hinged frames) that may, depend on the motor skills of the participant (Condy et al., 2021).
A game script was added to standardise the administration process and make the task more ‘game-like’. Game enjoyment was collated using a five-point Likert scale presented on a visual analogue scale (provided in Appendix B).

Similarly, the current research also utilised the validated measures of concept formation (i.e., the Matrix Reasoning and Similarities task in the WISC-IV battery) alongside the Alien Game. Given that language abilities could impact on performance (Alderson-Day & McGonigle-Chalmers, 2011), and it is not possible to minimise the language demands required on the Alien Game, the Similarities task was administered to establish verbal deficits and the impact on game performance. Hence, The Matrix Reasoning task was also administered to identify difficulties related to visuo-spatial skills.

Based on existing literature which posits parental occupation to be a predictor of performance on executive function tests (Last et al., 2018), parental occupation was also collected from the participants as part of the demographic information.

Finally, an objective measure of executive functions, The Childhood Inventory of Executive Function (CHEXI; Thorell & Nyberg, 2008) (provided in Appendix C), was completed by teachers to identify if problems associated with executive functions were also present in the academic setting.

1.11. Research Questions

The present study aimed to answer the following questions:

1. Is the updated Aliens Game a useful measure of concept formation in children aged 8-11 years?

2. Do typically developing children approach the test in a predictable, and exhibit a range of scores that do not have ‘floor’ or ‘ceiling’ effects?

3. Do demographic characteristics, existing measures of concept formation, or behaviour ratings of participants reflect their performance on the new game format?

2. METHODS

2.1. Epistemology

Ontology is a branch of philosophy that is concerned with the nature of ‘reality’, whereas epistemology is concerned with the ‘study of knowledge’ and ways of capturing and acquiring knowledge; for instance, it may involve questioning how knowledge can be generated and how that knowledge could be assessed to bring one's opinion or idea to real knowledge. Questions within the field of epistemology, may include: What is reality? If there is reality, when and how do we know this?

Within epistemology, there are different positions regarding what constitutes acceptable knowledge. The four main branches of research philosophy are positivism, interpretivism, pragmatism and realism. Positivist philosophy posits that only observable phenomena can produce facts and credible data, and thus the focus of this approach is on causality. On the contrary, the interpretivist position is concerned with capturing the subjective meanings and motivations of social phenomena to draw broad patterns using qualitative methods (Creswell & Clark, 2011). Whereas pragmatists (Kuhn, 1970) are concerned with utility and choose various practical “tools” to solve research questions. Finally, similar to positivism, realism, is also concerned with producing credible data and facts through observable phenomena; however, it does not make a clear cut between ‘true’ and false’; it considers that some aspects of reality remain hidden from observation (Duran, 2005). For this reason, while realists are guided by scientific criteria, they also recognize the importance of individual’s subjective experiences (Bhaskar, 1990).

This research adopts a critical realist position (Bhaskar, 1990; Greenwood, 1994). Given that one of the aims of a critical realist perspective is to understand entities, rather than merely describe them, it made sense to adopt this position in line with the research questions. One of the strengths of a critical realist position is that addresses the dualism between objectivity and subjectivism by acknowledging the differences between ontology (i.e., what is real) and epistemology (i.e., what we know). In other words, while it recognises the existence of an objective world, which

may exist independently of human cognition, it also recognises that knowledge is subjective, being bound by time and culture. In terms of its application to the current study, “executive function” is considered to be an entity that exists irrespective of individual’s background; however, “executive function” may be understood and conceptualised differently in different cultures. For example, although the inability to regulate one’s emotions and plan may be classified as signs of ADHD in the Western world, and lead to over-diagnosis in child and adolescent mental health services, these characteristics may be considered typical for a developing child in other parts of the world (e.g., in non-Western countries such as Turkey); (Timimi, 2017). Therefore, caution must be taken when interpreting the findings of this study as this research was conducted within a Western culture at a particular point in time, against a background ‘network’ of cognitive constructs.

2.2. Design

The study utilised a quantitative cross-sectional correlational design to address possible relationships between participants performance on the Alien Game and variables of interest: established measures of concept formation (i.e., WISC-IV Similarities and Matrix Reasoning) and objective measures of adaptive function (i.e., teacher-rated CHEXI). Through this research design, the researcher adopts an exploratory stance to evaluate the strategies employed by typically developing children, and their performance on the novel alien game measure. It was hoped that an exploratory approach to data analysis would enable the researcher to develop possible measures to find the best way to score the game.

Children’s scores on the Alien Game were compared with their scores on the Matrix Reasoning and Similarities tasks (WISC-IV), to establish concurrent validity. Teacher-rated CHEXI was used to assess executive functions in the classroom setting, which would provide predictive validity. Of note, although no norms has been established for the CHEXI, this instrument is believed to have good test-retest reliability (Thorell & Nyberg, 2008).
Finally, a 5-point Likert visual-analogue rating were used to explore the enjoyability of the Alien Game. A qualitative methodology was utilised to evaluate the feedback sought from participants regarding the Alien Game. Principles of thematic analysis

(Braun, 2008) were utilised to establish the common themes emerged from the feedback.

2.3. Recruitment

Participants were recruited from a mainstream primary school in London. Opportunity sampling was used as Local schools were contacted via e-mail and sent a copy of the research leaflet (provided in Appendix D) and 1 school expressed interest to participate. This was a small primary school, with one class of children for each of the year groups. This potentially limits the generalisability of this study since most state-funded schools in the UK comprise of bigger class sizes per year group and have very limited resources.

The inclusion criteria were: children aged 8- to 11-years who had sufficient conversational English to consent and participate in the study. As the study aimed to recruit a diverse sample of children, they were not excluded if they had neurodevelopmental conditions, such as ASD and ADHD. Given that executive functions are believed to be well established by 12 years of age (Best & Miller, 2010), this lower age range seemed appropriate for this research.

The exclusion criteria were: children who had insufficient English language, or severe developmental disorders (e.g., children with Down's syndrome with limited verbal skills), or severe visual or hearing impairments, such that they could not assent to take part or understand the test rubric or give a verbal response. Though they would not be excluded if they had corrected sight (e.g., wore glasses) and hearing (e.g., wore a hearing aid) nor any motor or sensory impairments.

2.4. Summary of Each Researcher's Contribution to Joint Project

This current study was completed by Pinar Marasli (the author), Emily Hay (EH) and Alexandros Bardis (AB). All three researchers were supervised by the same supervisor, and ethical approval was obtained jointly for the three research projects.

All three researchers were equally involved in all stages of the design and development of the Alien Game, and thus, the same set of Aliens were used by each researcher for data collection. Of note, Meabh Foley used a different Alien Game, with a new set of cartoon images comprising facial expressions specifically, to assess emotion recognition in primary school-aged children.

AB's study targeted children aged 6-7 years. AB presented the Alien Game in two different formats: Board condition versus Card condition. A 10-point cut-off was introduced for this group of younger children. AB aimed to find out whether: i.) game performance would differ between the Board versus Cards condition and the role of the Working Memory in performance; ii.) the Alien Game would correlate with established measures of concept formation; and iii.) it would be possible to conduct a General Linear Model (GLM) analysis, to see the influence of cultural factors (e.g., parental socioeconomic status) on game performance. AB developed a different scoring procedure for his own data and conducted a Spearman's Correlation to assess the relationship between game performance and established measures of concept formation. AB has not been able to conduct GLM analysis due to the small sample size (n=13). AB obtained game enjoyability ratings both for the Alien Game and the WISC-IV Similarities and Matrix Reasoning tasks, which allowed a descriptive comparison between the ratings to assess whether the Alien Game was perceived more favourably than the WISC-IV tasks.

The current author and EH were involved in the recruitment of participants, and both researchers tested the game on 18 children each (PM recruited children aged 8-9, and EH recruited children aged 10-11 years). EH's hypothesis were to find out whether a culturally fair test of concept formation could be derived from the Alien Game and to what extent the Alien Game would correlate with established measures of concept formation (i.e., WISC subtests) and executive functions (i.e., the CHEXI). PM's hypothesis were to examine whether the updated Alien Game is a useful measure of concept formation and whether the demographic characteristic, existing measures of concept formation, or behaviour ratings of participants reflect their performance on the new game format. It also aimed to find out whether the participants could exhibit a range of scores that do not have 'floor' or 'ceiling' effects.

PM and EH scored up their own data separately and afterwards checked each other's scores to correct any errors in the data entry. Following this, both researchers combined their data in order to have a bigger sample. Each researcher played an equal role in the development of the scoring procedure and conducted the data analysis together. The current author additionally obtained qualitative feedback on how the Alien Game could be developed further and conducted a thematic analysis for her study, while EH only obtained game enjoyment ratings for the Alien Game. Of note, the write-up of the thesis was carried out entirely separately by each researcher.

2.5. Development of The Alien Game

Pavitt (2017) engaged in the first phase of the Alien Game. The current study aimed to refine the Alien Game in order to overcome some of the limitations outlined by Pavitt. Some of the limitations of Pavitt's study were that not all Alien attributes had been named by the participants (e.g., the wings and the tails) and the study experienced a 'ceiling effect'. All attributes must be distinguishable to ensure that the game is, in fact, testing concept formation. Having distinguishable attributes is key to making the Alien Game a viable test and enabling examiners to identify those children who can reach the target quickly and accurately. Further, the current researchers wanted to even out the attributes to reduce to element of chance; and see if the new set of Alien would be better received by children. A game script was also added to make the game more standardised.

The researchers reviewed and redesigned all 24 pictures of the Alien used in Pavitt's (2017) study (given in Appendix E).

2.5.1. Game modifications

The following alterations were made to the game:

- The images were redesigned using Adobe Illustrator Software (version 26.4.1); (Adobe, 2022), to enhance the visual quality of the physical features and make them more distinct and prominent (e.g., the horns, wings, eyebrows). It was

hoped that these changes would make the game more Alien-like, and thus more appealing to younger children. See the grid in Appendix A for the Aliens created.

- The attributes of the Aliens were evened out, to minimise the risk of a ‘lucky guess’ from eliminating a large number of the items in the first question. Appendix F provides a comprehensive list of the attributes used in the design of the Aliens.
- Ambiguous and unclear attributes were removed (e.g., like the elephant trunk) and replaced with new attributes (e.g., a nose).
- The names of the Aliens were removed, to reduce the linguistic demands of the game.

Feedback from co-researchers was sought ahead of finalising the images. Based on the feedback given, further alterations were made to the Aliens to refine the images and produce a coherent set. For example, small attributes (e.g., the eyebrows) were refined to make them more vibrant and distinctive; the size of the attributes were standardised (e.g., eyes, arms, wings, horns) and repositioned, to make each of these attributes stand out. Shadows were added around the arms of the Aliens to make them more distinctive. For the nose, three dimensions were introduced: no nose, small nose, and big nose. Finally, Aliens with unintended attributes were removed (i.e., an Alien with a head and a body, see Appendix G) as these formed a superordinate attribute. Instead, it was decided to keep the overall shape of the Aliens as triangle, square and circle.

Once the images were finalised, a layout of the game was created. This enables the game to be accessible (e.g., cheaper) for clinical use, and reduces the motor demands associated with more intricate assessment methods (e.g., the use of a hinged board), which some children may find challenging. All 24 pictures of the Aliens were presented on a laminated A3 grid in four rows of six images (shown in Appendix A). The colours used in the game were chosen to make the game accessible to children who may be colour blind: the colours used in the game were restricted to contrasting shades of blue, yellow, and grey.

The children were permitted to play the game by asking as many questions as necessary until they identified the target Alien.

2.5.2. Game Script

The game rubric and instruction script were developed to standardise the game procedure for all participants as follows:

(a). 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 am thinking of".*

Play the game.

(b). If a participant asked a non-permitted question, reply: *"I can only answer yes or no."*

If they pause longer than 30 seconds before asking a question 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."*

Discontinue after 1 minute without a question.

2.6. Developing Measures to Score the Alien Game

The researcher also aimed to develop and simplify the scoring procedures outlined in Pavitt's (2017), to determine the best way to score the test.

2.6.1. Calculations Used to Score Game Measures

Similar to Pavitt's (2017) study, all measures were initially calculated by working at averages per trial (i.e., proportions). The strength of using averages was that it produced highly accurate scores, however, this method did not create enough variability in the data, and it produced decimals rather than integers, which are hard to calculate and interpret. In order to make the test more user friendly, the researcher calculated all measures by adding up the total numbers (i.e., actual numbers as absolute values). Using absolute values created more range in the data for clinical inspection. The data in the study was analysed using Statistical Package for the Social Sciences (SPSS) software Version 28 (IBM, 2021).

A brief explanation is provided below for the different metrics used:

- *Actual numbers* - added up absolute/actual values together.
- *Proportions* - calculated the proportion of scores from the number of Aliens remaining. For example, if 8 aliens were remaining, and a single question removed 2 Aliens, the minimum score for the minimum number of Aliens the question would remove would be 0.25 (2 divided by 8) and the maximum score for Aliens removed would be 0.75 (6 divided by 8).

2.6.2 Game Measures

The following measures were employed in the analysis to determine concept formation abilities and to establish what is the best way to score the test.

2.6.2.1 Total Questions (TQ). This is a new measure introduced in this study, derived by simply adding up all the questions asked in each trial.

2.6.2.2. Abstraction Score (Ab). As some children completed the trials by asking as few as three questions, the Ab was calculated based on the first three questions asked, on all four trials by each participant. The Ab was determined by adding up the minimum number of Aliens the questions could have removed (regardless of whether or not the question produced a correct answer). For instance, if the participant's first question was "Are you thinking of a yellow Alien?", the Ab score would be 7 whether or not the question produced a correct (n=7) or incorrect (n=17) response. This score was based on the AS developed in the DKEFS 20-Question test, and used in Pavitt's (2017) study; and was included in the current study to examine if this group of participants had similar abstraction scores.

2.6.2.3. Initial abstraction score (IA). IA was based on the D-KEFS scoring, and it consists of the minimum number of Aliens eliminated by the participant's first question across all four trials.

2.6.2.4. Ineffective Questions (IQ). This was determined by calculating all non-constraint seeking questions (e.g., not a yes/no question, fails to remove any items, repeated questions, incorrect guess attempt).

2.6.2.5. Hypothesis-seeking Questions (HS). HS score was determined by only adding up all the incorrect guess attempts (e.g., “is it Alien number 4”) whether or not the strategy was successful.

2.6.2.6. Weighted achievement score (WA). The WA was based on the scoring system used in DKEFS, which takes the total number of questions asked by individual participants in each trial, and compares that to the optimum number of questions, achieved in asking either too few or too many questions. When participants guessed the target Alien in fewer questions, they were given a low score because only lucky guesses could lead to correctly guessing the target Alien in 3 or fewer questions. The point scale was refined to ensure that participants were scoring proportionally to the total number of questions asked: fewer points are awarded if the participant guessed correct Alien in fewer than 4 or more than 7 questions. The WA point scale is shown in Table 2.

Table 2

Weighted Achievement (WA) Score Calculation

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

Abbreviations: TQ = total number of questions; WA = weighted achievement score of total questions

2.6.2.7. Weighted Ineffective Questions Score (WI) - Similar to the WA measure, a weighted scoring system was developed for ineffective questions, and each participant was assigned a new score based on their total ineffective questions score. The WI shows the total number of IQs asked by individual participants in each trial and allows the researcher to compare this to the optimum number of questions. Thus, each participant scored proportionality for the IQ's. Table 3 gives the weighted scoring scalar: more points were provided if participant's asked fewer ineffective questions.

Table 3

Weighted Ineffective (WI) Questions Score Calculation

Total IQ	WI score
0	5
1-2	4
3-4	3
5-6	2
7	1
8+	0

Abbreviations: IQ = total number of ineffective questions; WA = weighted achievement score of ineffective questions

2.6.2.8. Alien Game Score (AG). This new measure incorporates the WA and WI scores. So, it is a measure of total number of questions asked and the strategy the child is using. This is a single measure that is intended to capture overall game performance: high AG score indicates better performance.

2.6.2.9. Learning Slopes. Learning slopes were calculated for the TQ, Ab and IQ raw scores. Trial by trial comparisons were made to establish whether performance on the Alien Game improved across the four trials. This provides information regarding the experience of the game, and whether children's performance improved with practice.

2.7. Existing Measures

A number of established psychometric measures were used alongside the novel Alien Game measure, in order to address the concurrent and predictive validity of the novel measure. The Matrix Reasoning and Similarities subtests of the WISC-IV battery were used to measure concurrent validity, and the teacher-rated Childhood Executive Functioning Inventory (CHEXI) was used to test predictive validity.

2.7.1 WISC-IV Matrix Reasoning and Similarities

Matrix Reasoning is a non-verbal test based on Raven's progressive matrices tests and related instruments. This test is thought to measure a young person's visual perception as well as abstract reasoning. In this subtest, the children are presented with an array of pictures with one missing component; they are required to scan the pictures and select the option that fits the sequence from the options provided.

The WISC-IV Similarities task is a verbal test in which the examinee is asked to say how two items or ideas (e.g., a kiwi and melon) are related; a putative measure of verbal abstraction.

2.7.2. Childhood Executive Functioning Inventory (CHEXI)

Children's teachers were asked to complete the Childhood Executive Functioning Inventory (CHEXI; Thorrell & Nyber, 2008), to assess their 'real-world' adaptive function and behaviour. The CHEXI is available in different languages, and the English version has 24 items comprising four subscales: Working Memory (11 items, e.g., *"Has difficulty remembering lengthy instructions");* Inhibition (6 items, e.g., *"Has difficulty refraining from smiling or laughing in situations where it is inappropriate");*

Planning (4 items, e.g., *"Has difficulty with tasks that involve several steps");* and Regulation (5 items, e.g., *"Gets overly excited when something special is going to happen")*. The statements are rated on a scale from 1 to 5 (with "1" meaning definitely not true, and "5" meaning definitely true). The Working Memory domain is formed of the Working Memory and Planning subscales, while the Inhibition domain is formed of the Inhibition and Regulation subscales.

The CHEXI has been reported to have good reliability for both males and females. In a confirmatory factor analyses and cross-cultural validity study, it was found that the working memory and inhibition subscales had good psychometric properties, and are able to discriminate, with high sensitivity and specificity, between children who may have traits of ADHD, including within disparate cultures (Amukune & Józsa, 2021, Catale et al., Meulemans & Thorell, 2015)

Although formal normative data has not yet been established for the CHEXI, Catale et al. (2015) provided means, standard deviations and proposed cut-off scores based on the CHEXI scores of 242 typically developing children. These values are noted in Table 4.

Table 4

Combined means, standard deviations and suggested cut-off scores for the CHEXI Working memory and inhibition domains outlined by Catale et al. (2015, pg. 5)

Domain	Mean (SD)	Cut-off
Working Memory	24.05 (7.89)	34-35
Inhibition	25.65 (7.89)	32-34

2.8. Visual analogue scale and Qualitative Feedback

A 5-point Likert visual analogue scale (provided in Appendix B) was used to obtain feedback from the children after they completed the Alien Game, in order to address

the acceptability of the novel measure. Children were asked to either tick or circle the rating that best reflects their enjoyment of the game. The 5-point Likert scale ranged between “awful” to “fantastic”, with accompanying images.

Given the importance of developing a test that is suitable to use with young children, qualitative feedback was sought from them straight after the Alien Game. Children were asked “how could the Alien Game be improved?”, though it was made clear they did not have to provide feedback unless they wanted to. Their feedback was evaluated using simple thematic analysis.

In summary, feedback from the children were mostly positive. Children provided suggestions about how the game could be developed further, to make it more interesting and engaging for children of similar age. There seemed to be a common theme of increasing the diversity of Alien colours and attributes, as well as introducing rules to the game as a way of making the game more challenging. Responses obtained from the participants are provided in Table 5. Themes of initial codes (provided in Appendix H) and themes of final codes (provided in Appendix I) have been added to the appendices. It is important to be mindful of the possibility of social desirability bias when reviewing the feedback.

Table 5

Qualitative Feedback from Participants (N = 17)

Question	Participant Feedback
How could the Alien Game be improved?	• "It's really good…because…it's improving your guessing language…it's getting you good at understanding…it's funny." P1 • "I like guessing games like Guess Who?" P2 • "I wouldn't say it was really good or bad...If they could change the rules, like only if you could guess twice…or if you could only take a certain time to answer." P3 • "I liked it because guessing games are always fun…because they are like…you are a detective." P4 • "It was great…It would be much better if it had other things included.. like.. more colours…having a few lives.. so, one life is taken if you got the wrong alien.. it would be [also] good if there was one question you couldn't ask until the end…asking the colour makes it a lot easier, so [you] can make that [a rule] so you can't ask that until the end…or after 5 or 6 questions… so it can be a bit harder." P5 • "It was good." P6 • "You can add more colours. Each Alien could be a different colour. Like primary and secondary…and unique colours." P7 • "It was fun. I liked how it was." P8 • "You could add more colours and [features].. like a tongue sticking out… some could be sad or crying." P9 • "I liked it because it tested my memory." P10 • "I liked the game and how I was filling out…it amazed me… keep it as it is." P11 • "It was fun…maybe you could add other colours.. and like small spikey teeth…one reading a book…make them do signs with their hand, like a thumbs up." P12 • "Adding more Aliens would be better." P13 • "I liked how it had different shapes and colours. Keep it the same." P14 • "I liked it because.. it was challenging, I like challenging games…You can make it a little bit harder like…by...guessing two Aliens instead of one." P15 • "I liked that you had to use your brain to know which one it is…but [you could] have more aliens." P16 • "It was fun, and I like playing guessing games…you could add another colour or have one Alien with two colours, like a bi-colour." P17

2.9. Ethics

Ethical approval was obtained from University of East London Research Ethics Committee (provided in Appendix J). Initially, the SENCO and the Head Teacher of the school were contacted with the research details and flyer (provided in Appendix D) to ascertain whether they were interested to participate in the study. Once the Head Teacher agreed to their pupils being recruited, they were asked whether they wanted parents' to complete a consent form to opt-in their children into the study, or whether they were happy for parents to complete a parental consent opt-out form if they did not agree to their child taking part. It was explained that if the Head Teacher preferred the 'opt-out' option, they would be required to act in *loco parentis* (provided in Appendix K) in giving their consent for children to be included in the study. The Head Teacher indicated a preference for the 'opt-out' option. Following this, the Head Teacher was supplied with a copy of the organisation invitation letter (provided in Appendix L), Head Teacher's *loco parentis* Form (provided in Appendix K), Participants Information Sheet for Parents (provided in Appendix M), Participant Information Sheet for Children (provided in Appendix N) and Parental Consent Opt-Out Form (provided in Appendix O). A copy of the Parental Consent Opt-Out Forms were sent to parents by the school.

For those children whose parents had not opted them out, the researcher was given a list of children who were eligible for the study. Each child was approached individually by the researcher during school hours. Once the researcher and the child were in the testing room, the researcher read out the information sheet to each child and went through the consent form in full. Children were made aware that they could change their mind and stop testing at any point and that their information would be discarded. They were informed that, after completing the study, they had until the end of January 2023, to withdraw from the study, and if they withdrew, their data would not be included in the analysis. Children were told that their informed would be kept securely, and all identifiable information would be removed to maintain their confidentiality. Children were given an opportunity to ask questions before agreeing to take part in the study. If the child was happy to proceed, they were asked to complete the child consent form (provided in Appendix P). After they completed the tasks, children were read out the participant debrief sheet (provided in Appendix Q),

and asked whether they were negatively impacted by taking part. No deception or hazardous procedures were used, and no children were negatively affected. Children were provided with the researcher's and universities contact details in case they wanted to raise questions or concerns regarding the study.

2.10. Materials

The following materials were used in this study:

- A3 laminated Alien Game 6x4 grid
- Standardised instruction script and response sheets
- Opt-out forms
- Information sheets
- Consent forms
- Debriefing sheet
- Demographics sheet
- A pen and paper
- Clipboard
- A table and chair
- The CHEXI (teacher-rated)
- 5-point Likert visual analogue scale
- Matrix Reasoning stimulus booklet

2.11. Procedure

Parent opt-out forms were sent to parents by the school. A list of potential participants names were produced by the participating school's SENCO for those children's parents who had not opted-out their child to participate in the study. The researcher randomly selected potential participants from the list provided.

The Alien Game was played first. The researcher read out the standardised game instruction script (see section 2.5.2.) and checked if the child had any questions before proceeding to the game. Alien Game was then presented on a laminated A3

(6x4) grid, approximately 10cm away from the child. The grid is presented in Figure 2 and provided in Appendix A.

Figure 2. Alien Game on a 6x4 Grid

Each child played four rounds of the Alien Game. The order of the target Aliens were kept the same to ensure consistency across the trials. The target Alien's were in the following order: Alien number 16, 8, 17 and 24, as shown in Figure 3.

Alien No.16 **Trial A**

Alien No.8 **Trial B**

Alien No.17 **Trial C**

Alien No.24 **Trial D**

Figure 3. Order of Target Alien's Across

Children's questions for each trial were recorded on the response sheet. At the end of the fourth trial, each child was asked to rate their game enjoyment using the visual analogue scale (provided in Appendix B). Then, the children were given the option to provide verbal feedback on the game.

Next, the Similarities and then the Matrix Reasoning tests were administered using Satz-Mogel's method (Satz & Mogel, 1962). The instructions were read out from the stimulus book and the participants were allowed enough time to complete each question. Following completion of the tasks, the child was returned to their classroom and their class teacher was asked to complete the CHEXI.

The test took between 25-45 minutes to complete with each child. Children were reminded that they could take breaks in between tasks, to minimise boredom and fatigue. They were also offered a cup of water between tasks. None of the children experienced distress as a result of participating in the study.

2.12. Confidentiality

All information relating to the child's demographics and responses were recorded on a record form (provided in Appendix R-S). Each child was given a code (and thus no names were recorded) and the same code was written on all corresponding documents. Each child's details and responses/scores were then transferred on Excel and SPSS databases, respectively. These were kept on a password protected Windows laptop using the secure OneDrive associated with the researcher's UEL IT account. Paper documents were stored in a secure and locked cabinet, and were securely shredded after the date to withdraw from the study had passed.

2.13. Participants

2.11.1 Demographics

Participants in this study were 33 typically developing children. As shown in Table 6, the sample consisted of 16 males and 17 females, ranging in age from 100 to 132 months (Mean = 9.8 years, SD = 0.98). Participants were from a range of ethnicities,

with the three predominant groups being Western European (75.8%), non-Western European (6.1%), and Other (18.2). Of the 33 participants,13 identified speaking English as their primary language (EPL), and 20 speaking English as an additional language (EAL). None of the participants reported having any visual or motor needs. Participants were from a range of socio-economic backgrounds (based on reported parental occupation grade). Of the 33 participants, 1 participant's parent had a higher professional occupation (grade 5), 6 parents had an intermediate level occupation (grade 4), 7 parents were in the small employers and own account category (grade 3), 12 had lower supervisory and technical occupations (grade 2), and 1 had a semi-routine and routine occupation (grade 1). 6 children were unable to recall their parents' occupation, and this was recorded as missing data.

Table 6

Demographic Characteristics of Participants

Demographic variable	Summary statistic	
Gender, *n (%)*		
Male	16	(48.5%)
Female	17	(51.5%)
Age, *mean (SD)*		
Age in Months	117. 66	(11.73)
Age in Years	9.8	(0.98)
Ethnicity group, *n (%)*		
Western European	25	(75.8%)
non-Western European	2	(6.1%)
Other	6	(18.2)
English as Primary Language, *n (%)*		
Yes	13	(39.4%)
No	20	(60.6%)
Socio-economic grade*, *n (%)*		
5= Higher managerial/professional occupations	1	(3.0%)
4= Intermediate occupations	6	(18.2%)
3= Small employers and own account workers	7	(21.2%)
2= Lower supervisory and technical occupations	12	(36.4%)
1= Semi-routine and routine occupations	1	(3.0%)
Missing data	6	(18.2%)

Note. * The socio-economic grading system descriptors were taken from *The National Statistics Socio-economics classification*

2.13.2. Sample Characteristics

Descriptive statistics for the participants' scores on the WISC-IV Similarities and Matrix Reasoning tasks are presented in Table 7. A breakdown has been provided for the sex and language groups, English as primary language (EPL) or English as additional language (EAL), as well as for the overall sample. Descriptive statistics are provided for the teacher-rated CHEXI scores, presented in Table 7. Raw scores were transformed into age-related scaled scores. The population normative data (Mean = 10, SD = 3) was used to compare this sample of children's age-scaled scores across the various tests.

In the WISC-IV tasks, females and males scores in the Similarities were overall high average, which is above the population norms. For the Matrix Reasoning, both males and females, scored in the average range, overall. Children in the EPL and EAL groups scored similarly across the Similarities and Matrix Reasoning tasks, though the scores for the EPL group were slightly higher for both tasks. Scores overall confirm that children in this sample performed somewhat better in Similarities than in Matrix Reasoning.

As the sample number was small, and most of the variables were ordinal rather than interval-level, non-parametric analyses were used to address group differences. Mann-Whitney U tests with exact tests (based on the sampling procedures available in SPSS) were conducted to establish whether there are any group differences. The results indicate that there were no differences between scaled scores by sex nor language groups. Descriptive statistics and Mann-Whitney U test values are presented in Table 8.

Table 7

Participants' WISC-IV Similarities and Matrix Reasoning Age-scaled Score

Demographic		N	WISC-IV		Mann-Whitney U	Z	Exact Sig. (2-tailed)
			Similarities	Matrix Reasoning			
Sex	Male	16	13.00 (2.80)	10.25 (2.52)			
	Female	17	12.53 (2.27)	10.88 (1.96)			
	Overall Sample	33	12.76 (2.51)	10.57 (2.24)			
	Group Comparison Similarities	33			119.00	-.619	.547
	Group Comparison Matrix Reasoning	33			105.00	-1.133	.265
Language							
	EPL	13	13.23 (3.06)	11.07 (2.01)			
	EAL	20	12.45 (2.11)	10.35 (2.39)			
	Overall sample	33	12.76 (2.51)	10.63 (2.27)			
	Group Comparison EPL	33			114.00	-.596	.562
	Group Comparison EAL	33			123.00	-.262	.805

Abbreviations: EPL = English as primary language; EAL = English as additional language

For the CHEXI, the scaled scores for the Working Memory and Inhibition domains based on the Catale et al. (2015) data parameters are shown in Table 8. Females Inhibition and Working Memory scaled scores were higher than males, with females

scoring in the above average range for both domains. Children with EPL and EAL scored similarly for the Working Memory domain; while for Inhibition, children with EAL scored slightly higher. The overall scores show the children in this sample scored in the average range for both domains, though their Inhibition scores were slightly higher than the population average. The Mann-Whitney U test showed that there were no group differences (i.e., sex and language) in these measures. These values are presented in Table 8.

Table 8

CHEXI Age-scaled Scores for the Working Memory and Inhibition Domains

Demographic		N	CHEXI Working Memory	CHEXI Inhibition	Mann-Whitney U	Z	Exact Sig. (2-tailed)
Sex	Male	16	10.37 (3.22)	11.18 (2.76)			
	Female	17	11.41 (2.39)	12.70 (2.82)			
	Overall Sample	33	10.90 (2.83)	11.97 (2.85)			
	Group Comparison Working Memory	33			107.00	-1.06	.296
	Group Comparison Inhibition	33			95.50	-1.47	.14
Language							
	EPL	13	11.07 (2.39)	11.77 (2.55)			
	EAL	20	10.80 (3.13)	12.10 (3.09)			
	Overall Sample	33	10.90 (2.83)	11.96 (2.85)			
	Group Comparison Working Memory	33			127.00	-.11	.92
	Group Comparison Inhibition	33			123.00	-.24	.82

Abbreviations: EPL = English as primary language; EAL = English as additional language

Given that neither sex nor having English as a primary language has influenced scores on their WISC And CHEXI measures, this warrants us to treat sex and language as separate variables in subsequent analysis. As mentioned above, this sample of children were a verbally able group compared to the norms, which may limit the generalisability of this study.

3.RESULTS

3.1. Methods of Analysis

3.1.1 Exploratory Data Analysis

An exploratory data analysis (EDA) was carried out to check for data distribution, to identify outlying scores, and correct any errors. All values were double checked and verified by a co-researcher working on the data set. Due to the small sample size, Shapiro-Wilks test was performed to check for normality of distributions: A measure was considered non-normal if it had a significant value (i.e., the hypothesis of normality was rejected if it had a significance value of $p < .05$). Histograms were reviewed for the shape of the distribution, and boxplots were checked for any outlier; the skewness (<1), kurtosis (<3) parameters were also used to assess normality of distributions.

The EDA showed that most of the variables were not in a normal distribution, and accordingly, non-parametric tests were performed for all subsequent analysis. Spearman's rank correlations were used to address the relationships between the Alien Game scores and the established measures. Wilcoxon's Signed-Rank tests were used to determine whether there were any changes in performance across the trials, which also served as a preliminary reliability analysis. Cohen's effect size statistics were used to interpret the effect sizes: values between .10-.29 indicated a small effect; .30-.49 indicated a moderate effect; and values greater than .50 indicated a large effect size. Mann-Whitney U tests were also performed to compare scores on the Alien Game and established measures between the sex and language groups. The Likert-ratings of game enjoyment were analysed by looking at the frequencies of responses.

Of originally 34 participants who participated in the study, one child's scores were excluded from the data analysis as it was difficult to make sense of their game strategy and they were reliably an outlier on all game measures. This will be

discussed further below.

3.1.2. Test of Normality

Table 11-12 presents the descriptive statistics both for the measures explored for the Alien Game, as well as for the CHEXI and the Similarities and Matrix Reasoning tasks. The measures that did not meet the normality assumption are indicated in bold in Table 11-12, and described below.

3.1.2.1. Alien Game measures

- *Ab score using proportions* - Initially, a minimum (min) and maximum (max) Ab scores were calculated. The Ab (Min) captured the minimum number of Alien's that could be eliminated by the first three questions across the 4 trials, and the Ab (Max) captured the maximum number of Alien's that could be eliminated by the first three questions (from the remaining number of options). Next, an Ab difference (Diff) score was calculated by subtracting the Ab (Max) from the Ab (Min). These were first calculated using proportions, and using absolute values.

  However, descriptive statistics showed that the means were low, had a small standard deviation, and the range of scores were very small, which made interpretation of the data difficult. It also created decimals, ranging from 0 to 1.40, reducing user friendliness. Using proportions also meant that some children obtained scores of 0 for the IQ's asked (e.g., questions that failed to eliminate any Aliens), which reduced the variability in the data. For example, if 4 Aliens were remaining, and the subsequent question failed to remove any aliens (i.e., 0 Aliens), the Ab (min) score for this question was scored as zero (i.e., 0 divided by 4). In this instance, the maximum number of Aliens that could be removed is 4. The Ab (max) was calculated by dividing 4 (i.e., the maximum number of Aliens that could be removed) by 4 (the number of remaining aliens), which gave a score of 1. The Ab (Diff) was calculated by subtracting the Ab (max) score from the Ab (min) score. In this case, it would

be 1 minus 0, which would give an Ab (Diff) score of 1.

- *Ab score using absolute values* - The Ab (Min), Ab (Max) and Ab (Diff) scores were recalculated using absolute values. Once again, Ab (Min) gave the best range of scores. Adding the numbers together created a bigger range (45 to 76) in the scores, which made the interpretation easier and so increased the 'user friendliness' of scoring.

  Accordingly, it was decided that using absolute values was the best way of calculating the Ab component of performance.

- *Ab score for all of the questions versus the first three questions asked* - Ab score for all questions for each trial was calculated, using actual numbers to derive the Ab (ALL) measure. This Ab (ALL) measure had the advantage of incorporating all of the questions asked, and could potentially be a better measure than the Ab (Min), which only included the first three questions.

  The data analysis showed that, there was not much variability in the data and it had a very small range, and that the means and the standard deviations were very small for these scores. In clinical practice, these would be hard to calculate and interpret.

- *Ab (Diff) for all questions versus Ab (Diff) for the first three questions* - The difference scores were calculated for the Ab in various ways. Firstly, an average score of the Ab (Diff) scores were calculated for each participant, and each trial. For example, if the participant reached the target Alien in 8 questions, the Ab (diff) score for each of the 8 questions were added together and divided by 8. Once this score was calculated for all participants across the 4 trials, this sum was then divided by 4 to derive the Ab (Diff) for all questions. This was as a sum and averaged.

  None of these measures produced normally distributed data, and there was little variability in the data, for both proportions and total scores. The Ab (Diff)

yielded no values of any size and was also the most complex way of scoring performance.

This data analysis confirmed that Ab (Min), total score was the best way to calculate this component, though it might not work for every child. It also seemed to be a good test to make out of as it exhibited no floor or ceiling effects.

- *IA* - IA is based on the first question asked in each trial. The researcher tried various ways to calculate the IA, however when the IA was derived using both proportions (the first question of each trial added together and divided by 4) or with actual numbers (the first question of each trial added together), none of the values produced good range of data. The data showed a Leptokurtic distribution, with most scores clustered in the centre of the distribution, meaning that most children had obtained the same score. This was due to the majority of the children in this sample removing a large fraction of non-targets (i.e., 8 out of 24 Aliens) in their first question. Few children asked hypothesis-seeking questions (i.e., “Is it Alien number 10”) in the first question. The IA was therefore not deemed to be a useful measure of game performance.

- *HS* -The researcher also considered whether it would be worth looking at hypothesis-seeking questions (HS) versus constraint-seeking questions. However, since the great majority of questions were constraint-seeking, only a score for the HS were calculated.

  However, given that there were not many HS overall, the HS was instead added to the calculation of IQ’s (i.e., questions that failed to remove any Aliens), which is explained further below.

- *TQ* - The TQ was calculated using actual numbers, being the total number of questions asked across the 4 trials. This simple way to calculate the TQ score also gave the biggest range of scores, (between 14 and 49) using actual numbers is, once again, better for practical use and interpretation.

- *WA* - For the WA, the researcher created a different scoring method to the one developed by Pavitt (2017), to reduce the element of chance associated with guessing the target Alien. Initially, the WA scale was based on the proportion of TQ asked and averaged across the 4 trials. However, as it was decided that using absolute values to calculate the TQ score was best, the lower end of the WA scale was adjusted accordingly. Absolute values of TQ made the WA score easier to calculate, and increased the range (between 3 to 19) of scores, standard deviation (4.81) and the mean (12). A histogram indicated that most children were clustered around the top scores, meaning that they showed good performance in the game. The benefit of this measure is that it makes transparent the children who perform poorly, as they would appear in the lower 'tail' of the distribution.

  When participants guessed the target Alien in fewer questions, they were given a low score because only lucky guesses could lead to correctly guessing the target Alien in 3 or fewer questions. Therefore, each participant scored proportionally using the WA scale. Participants were given higher points if they guessed the correct Alien between 4 and 6 questions as this was considered to show good concept formation skills. The WA scoring scalar is presented below in Table 9.

Table 9

Weighted Achievement (WA) Score Calculation

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

Abbreviations: TQ = total number of questions; WA = weighted achievement score of total questions

- *IQ* - The IQ comprised questions that failed to eliminate any items, including repeated questions; not a yes/no question; incorrect guesses (i.e., HS); any other questions that failed to remove targets. The IQ histogram showed a good range in scores (ranging from 1 to 31), with no outliers, in a normal distribution.

- *WI* - A weighted scoring scalar was developed for ineffective questions, based on the total number of ineffective questions asked in each trial. This allows each participant to score proportionality and for the examiner to compare this score to the optimum number of questions. This score showed a normally distributed data, and produced good variability in the scores, ranging from 0 to 15 and the boxplot showed no outliers. Table 10 gives the weighted scoring scalar: more points were provided if participant's asked fewer ineffective questions.

Table 10

Weighted Ineffective Questions (WI) Score Calculation

Total IQ	WI score
0	5
1-2	4
3-4	3
5-6	2
7	1
8+	0

Abbreviations: IQ = total number of ineffective questions; WA = weighted achievement score of ineffective questions

- *AG-* The researcher also aimed to identify a useful single score to indicate the participants overall game performance: the Alien Game overall score. Given that the WA and WI scores had proved to be good measures of game performance, it made sense to incorporate these scores into a composite measure of test performance. So, this measure would identify the children who performed well (i.e., asked the fewest questions, and the fewest ineffective questions) and children who struggled (i.e., who asked the greatest number of questions, and asked ineffective questions).

  The Alien Game score was calculated using absolute values (i.e., numbers added together), which produced good range in the data, and where a lower number indicated poor performance, and a higher number indicated better performance, which would be easy to report and interpret.

3.2.3. Strengths and Limitations of Using Absolute Values

Using proportions produced very accurate ways of calculating game performance, giving the closest representation of performance per trial. However, they did not yield enough variability in the data (because most children obtained a score of 0 or close to either for their abstraction score or because they failed to eliminate alien). Using proportions is also an intricate method to approach scoring, which may pose challenges to professionals who may struggle with the calculations required.

The advantage of using the absolute values is that this provides a more straightforward way to calculate the scores, by teachers and other professionals, and to get a better sense of performance. Using the absolute values produces countable scores, which make more sense to teachers and clinicians. However, we should be mindful that average scores (i.e., calculated using proportions) would not be influenced by number of trials completed; which might be useful where there is variability in number of trials completed (e.g., due to disengagement or fatigue). This was not a concern for this study, as all of the participants completed all four trials.

Table 11

Descriptive Statistics for Initial Alien Game Measures

		Mean	SD	Min.	Max.	Range
Alien Game Measures						
IA	IA (AD)	32.12	5.75	11.00	48.00	37.00
	IA (AAV)	8.03	1.44	2.75	12.00	9.25
	IA (PA)	1.33	0.26	0.46	2.00	1.54
	IA (PAV)	0.33	0.06	0.11	0.50	0.39
Ab	Ab Min (AD)	65.15	7.07	45.00	76.00	31.00
	Ab Min (AV)	16.29	1.77	11.25	19.00	7.75
	Ab Min (PA)	4.32	0.55	2.45	5.21	2.76
	Ab Min (PAV)	1.06	0.21	0.05	1.30	1.25
	Ab All (PA)	1.19	0.26	0.66	1.62	0.96
	Ab All (PAV)	0.30	0.06	0.16	0.40	0.24
	Ab Diff (AA)	0.51	0.26	0.18	1.16	0.98
	Ab Diff (AAV)	0.13	0.06	0.05	0.29	0.08
	Ab Diff of Min scores (added)	2.55	1.00	1.33	6.10	4.77
	Ab Diff of Min scores (averaged)	0.64	0.25	0.33	1.52	1.19
TQ and WA	TQ (added)	28.21	8.34	14.00	49.00	35.00
	TQ (average)	7.05	2.08	3.50	12.25	8.75
	WA	12.06	4.81	3.00	19.00	16.00
HS	HS (added)	3.36	3.44	0.00	11.00	11.00
	HS (average)	0.84	0.86	0.00	2.75	2.75

Abbreviations: IA = initial abstraction; Ab = abstraction Score; TQ = total questions; WA = total questions weighted achievement score; HS= hypothesis seeking; AD =

actual added; AV = actual averaged; AAV = all averaged; PA = proportion added; PAV = proportion averaged; AA = all added; Diff = difference.

Table 12

Descriptive Statistics for Initial Alien Game Measures

		Mean	SD	Min.	Max.	Range
Alien Game Measures						
IQ	IQ (proportion)	2.51	1.94	0.25	6.50	6.25
	IQ (ACO)	11.91	8.42	1.00	31.00	30.00
	IQ (ACinHS)	13.39	9.56	1.00	32.00	21.00
WI	WI	7.48	4.55	0.00	15.00	15.00
AG	AG (averaged)	1.44	2.18	-3.50	4.50	8.00
	AG (added)	19.55	9.03	3.00	34.00	31.00

Abbreviations: IQ = ineffective questions; WI = weighted ineffective questions score; AG = alien game score; ACO = actual values counted once excluding HS; ACinHS = actual values counted once including HS.

3.2.4. Established Measures

As mentioned above, all raw and scaled scores were checked by a co-researcher and errors were corrected. WISC-IV Similarities and Matrix Reasoning scaled scores showed a normally distributed data. Box plot did not show any outliers for the Similarities scaled score. However, Matrix Reasoning had two outliers, with one participant scoring extremely poorly and another scoring exceptionally high.

The CHEXI Working Memory domain showed a non-normally distributed data, with a long tail (i.e., scores clustered in the higher scores), however, this distribution is typical of symptoms scales. Further, the mean (10.91) is in the average range and there seems to be a good range in scores.

3.2.5. Final Measures of Game Performance

At the end of the exploratory data analysis, the best measures to use for game performance were established. These measures seemed to have no floor or ceiling scores, had a broad range of scores, and were calculated using absolute values rather than proportions. The final measures are mentioned below:

- TQ
- IQ (inc. hypothesis-seeking questions)
- Ab (Min)
- WA
- WI
- AG (sum of WA and WI scores)

3.3. Associations within Alien Game Scores

Table 13 gives a summary of Spearman's rank correlation coefficients, within the six Alien Game measures; moderate and large effect sizes are indicated in bold.

When comparing variables within the same measure, there was a large correlation between *TQ* and IQ: as the total number of questions asked goes up, so does the number of ineffective questions asked. There are large negative correlations between *TQ* and WA and AG as expected. There appears to be a moderate positive correlation between *TQ and Ab:* as the total number of questions asked goes down, the Ab score increases.

However, there were no correlations between Ab and any of the other game measures.

There are large negative correlations between IQ and WA, IQ and WI, and IQ and AG.

There are large positive correlations between *WA* and *WI*, and *WA* and *AG.*

Finally, there is a large positive correlation between *WI* and *AG*.

In this typically developing sample, TQ, IQ, WA, WI, and AG appear to be highly related measures, presumably reflecting the same cognitive capacities. Ab, in contrast, seems to be independent and potentially captures a different underpinning function.

Table 13

Spearman's Rank Correlation Coefficients between Alien Game Measures

Measure	TQ	Ab	IQ	WA	WI	AG
TQ	1.00					
Ab	**.36***					
IQ	**.94****	.23				
WA	**-.89****	-.16	**-.89****			
WI	**-.90****	-.23	**-.96****	**.84**		
AG	**-.92****	-.21	**-.96****	**.94**	**.96**	1.00

Note. Significant values are indicated in **bold**

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Abbreviations: TQ= total questions; Ab= abstraction score; IQ= ineffective questions; WA = total questions weighted achievement score; WI= ineffective questions weighted achievement score; AG= alien game score.

3.4. Concurrent Validity

Concurrent validity was addressed by comparing Alien Game performance with the WISC-IV Similarities and Matrix Reasoning scores, using Spearman rank

correlations, which can be seen in Table 14. Moderate and large effect sizes are indicated in bold.

No associations were found between the Alien Game and Similarities. *TQ and* IQ were negatively associated the Matrix Reasoning.

3.5. Predictive Validity

A summary of Spearman's rank correlation coefficients between the Alien Game measures and the *CHEXI* subscales is also provided in Table 14. CHEXI Working Memory and Inhibition domains were not associated with any of the Alien Game measures.

Table 14

Spearman's Rank Correlation Coefficients Between Alien Game and Other Measures

Game Measure	Concurrent Validity		Predictive Validity	
	WISC-IV Similarities	WISC-IV Matrix Reasoning	CHEXI Working Memory	CHEXI Inhibition
TQ	.00	**-.35***	-.07	.12
Ab	.08	-.16	.15	.17
IQ	-.05	**-.34**	-.17	.01
WA	-.03	.25	.09	-.06
WI	.10	**.33**	.16	-.07
AG	.04	.29	.16	-.05

Note. Significant values are indicated in **bold**

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Abbreviations: TQ= total questions; Ab= abstraction score; IQ= ineffective questions; WA = total questions weighted achievement score; WI= ineffective questions weighted achievement score; AG= alien game score.

3.6. Influence of demographic data on Alien Game Measures

3.6.1. Alien Game Measures and Sex

A Mann-Whitney analysis was conducted to establish whether Sex had an influence on Alien Game Performance. The analysis indicated that there were no sex differences for any of the Alien Game Measures (i.e., TQ, Ab, IQ, WA, WI and AG). The values are provided in Table 15.

Table 15

Mann-Whitney U Test for Alien Game Measures by Sex

	Descriptive Statistics			Test Statistics		
Measure	Sex	Mean	SD	Mann-Whitney U	Z	Exact Sig. (2-tailed)
TQ	Male	29.81	8.92	111	-.90	.38
	Female	26.71	7.72			
Ab	Male	64.44	7.63	122	-.51	.62
	Female	65.82	6.65			
IQ	Male	13.06	8.84	114	-.78	.45
	Female	10.82	8.12			
WA	Male	11.44	4.95	116	-.72	.48
	Female	12.65	4.76			
WI	Male	6.38	4.56	96	-1.43	.16
	Female	8.53	4.42			
AG	Male	17.81	9.25	108	-1.01	.32
	Female	21.18	8.79			

Note. Correlation is significant at the 0.05 level (2-tailed).

TQ and IQ: A lower mean indicates better performance. Ab, WA, WI and AG: A higher mean indicates better performance.

Abbreviations: TQ= total questions; Ab= abstraction score; IQ= ineffective questions; WA = total questions weighted achievement score; WI= ineffective questions weighted achievement score; AG= alien game score; SD= standard deviation.

3.6.2. Alien Game Measures and Language

A Mann-Whitney U Test was also conducted for language groups (English as Primary Language), to see whether language had an impact on scores on the Alien Game Measures. The analysis indicated that there were no group differences in any of the Alien Game Measures (i.e., TQ, Ab, IQ, WA, WI and AG). The values are provided in Table 16.

Table 16

Mann-Whitney U Test for Alien Game Measures by Language

	Descriptive Statistics			Test Statistics		
Measure	Language	Mean	SD	Mann-Whitney U	Z	Exact Sig. (2-tailed)
TQ	EPL	27.08	8.14	110	-.74	.47
	EAL	28.95	8.59			
Ab	EPL	64.38	8.20	129	-.04	.98
	EAL	65.65	6.40			
IQ	EPL	10.46	8.18	107	-.87	.40
	EAL	12.85	8.64			
WA	EPL	12.77	5.54	110	-.74	.47
	EAL	11.60	4.37			
WI	EPL	8.00	5.12	114	-.59	.57
	EAL	7.15	4.25			
AG	EPL	20.77	10.42	110	-.72	.48
	EAL	18.75	8.20			

Note. Correlation is significant at the 0.05 level (2-tailed).

TQ and IQ: lower mean indicates better performance.

Ab, WA, WI and AG: higher mean indicates better performance.

Abbreviations: TQ= total questions; Ab= abstraction score; IQ= ineffective questions; WA = total questions weighted achievement score; WI= ineffective questions weighted achievement score; AG= alien game score; SD= standard deviation; EPL= English as primary language; EAL = English as additional language.

3.8. Reliability Analysis

Wilcoxon Signed-Rank test (repeated measures) was used to see if there were any differences between the scores for TQ, Ab and IQ over the four test trials. Wilcoxon Signed-Rank test confirmed that there were few substantial differences for TQ, Ab and IQ learning slopes; however, there were no reliable improvements or deteriorations from Trial A to Trial D. Trials with significant differences are reported and indicated in bold in Table 17.

Table 17

Descriptive Statistics and Wilcoxon Signed-Ranks Test for Learning Slopes (TQ, Ab and IQ)

		Descriptive Statistics				Test Statistics	
Measure	Trials	Mean	SD	Min.	Max.	Z	Exact Sig (2-tailed)
TQ							
	Trial A	6.64	*3.00*	3.00	17.00		
	Trial D	7.91	3.47	4.00	18.00		
	Trial A – Trial D					-1.80	.07
Ab							
	Trial A	15.91	4.21	3.00	21.00		
	Trial D	17.82	2.24	12.00	20.00		
	Trial A – Trial D					**-2.50**	**.01**
IQ							
	Trial A	2.85	3.11	.00	12.00		
	Trial D	3.73	4.79	.00	21.00		
	Trial A – Trial D					-.36	.72

Note. Correlation is significant at the 0.05 level (2-tailed). Significant values are indicated in **bold.** TQ and IQ: the lower the mean the better the performance. Ab: the higher the mean the better the performance.

Abbreviations: TQ= total questions; Ab= abstraction score; IQ= ineffective questions; WA = total questions weighted achievement score; WI= ineffective questions weighted achievement score; AG= alien game score; SD= standard deviation; EPL= English as Primary Language; EAL = English as Additional Language.

3.9. Game Acceptability and Feedback

3.9.1 Alien Game Feedback

The feedback on the 5-point Likert scale regarding game enjoyment was assessed by looking at the frequency of ratings. Of the 33 participants, 10 respondents rated the game “fantastic”, 21 rated it as “really good”, and 2 rated it as “okay”. The frequency of the ratings are provided in Table 18.

Table 18

Participants Game Enjoyment Ratings

Game Rating	No. of participants	Percentage
Fantastic	10	30%
Really Good	21	64%
Okay	2	6%
Not Very Good	0	0%
Awful	0	0%

3.9.2 Evaluation of Game Attributes

All 13 Alien attributes were asked about by participants across the four trials. Table 19 shows the frequency of attributes asked across the four trials.

Twenty-six percent of the questions were about the colour of the Alien. Fourteen percent were about the eyes and the legs/feet of the Aliens. Wings were asked nine percent of the time, and eyebrows were asked six percent of the time. Five percent of the questions were about the tail, antenna, nose and horn attributes. Four percent were about the shape and the teeth attributes. Two percent of the questions were regarding the ears, and one percent was regarding the arms/hands. Of note, various names were used for the antenna (such as “planty things”), horns (such as “pointy things”) and nose (such as “koala nose”).

Table 19

Summary of Alien Attributes

	Frequency	
Alien Attributes	No. of times asked	Percentage
Colour	200	26%
Eye	107	14%
Legs/Feet	107	14%
Wings	72	9%
Eyebrows	45	6%
Tails	40	5%
Antenna	37	5%
Nose	37	5%
Horns	36	5%
Shape	31	4%
Teeth	31	4%
Ears	11	2%
Arms/Hands	10	1%

3.9.3. Game Feedback

Seventeen of the 33 participants provided feedback about “how the Alien Game be improved?”. These five (29%) stated that “nothing” needed to be changed, while 12 participants provided suggestions regarding how the Alien Game could be improved. Suggestions were analysed by the researcher using thematic analysis. Table 20 provides a summary of these suggestions using variable codes, as well as the frequency and percentage of the responses.

Five participants (29%) stated that the Alien Game could be designed with new/additional colours (e.g., creating some Aliens with multi-colours). Three participants (18%) highlighted that the game could be made more challenging (e.g., by limiting the number of guesses allowed per trial; not being able to ask about the colour of the Alien; and increasing the number of target Aliens per trial). Two participants (12%) said that more Aliens could be added to the grid. Finally, two

participants (12%) thought new attributes could be introduced, such as facial expressions or hand gestures, to make the game more interesting. Table 20 provides a summary of the feedback. Direct quotes from the participants are also outlined below. The suggestions from participants are in line with the search strategy adopted during the game. For instance, most participants asked about the colour of the Aliens as early as the first question across the four trials. Hence, feedback from participants suggests that either introducing more colours or restricting the examinee from asking about the colour of the Aliens as their initial question could potentially make the game more interesting and challenging.

Table 20

Summary of Thematic Analysis Codes and Frequency of Feedback

Codes	No. of participants	Percentage
No Changes	5	29%
Introduce New Colours	5	29%
Increase Game Difficulty	3	18%
Add More Aliens	2	12%
Introduce More Attributes	2	12%

3.9.4. Themes with Quotes from Participants

3.9.4.1 Theme 1: No Changes

Five participants commented that they thoroughly enjoyed the game in its current form. They did not feel the game needed changing.

> "It's really good…because it's improving your guessing language.. It's getting you good at understanding…it's funny." P1
>
> "It was fun. I liked how it was." P8
>
> ""I liked the game and how I was filling out…it amazed me… keep it as it is." P11

"I liked how it had different shapes and colours. Keep it the same." P14

3.9.4.2. Theme 2: Introduce New Colours

An important theme was the feedback about the colours of the Alien. Participants suggested introducing more variation to the colours, such as having Alien's with bi-colours.

"You could add another colour or have one Alien with two colours, like a bi-colour." P17
"You could add more colours." P9

3.9.4.3. Theme 3: Increase Game Difficulty

Some participants indicated that the rules of the game could be made more challenging. Ideas were offered, such as limiting the number of incorrect guesses.

"If they could change the rules, like only if you could guess twice…or if you could only take a certain time to answer" P3
"Asking the colour makes it a lot easier, so [you] can make that [a rule] so you can't ask that until the end of after 5 or 6 questions… so it can be a bit harder." P5
You can make it a little bit harder like…by..guessing two Aliens instead of one." P15

3.9.4.4. Theme 4: Add More Aliens

Participants mentioned that more Aliens could be added to the Alien Game grid.

"Adding more Aliens would be better." P13
"[You could] have more Aliens." P16

3.9.4.5. Theme 5: Introduce More Attributes

Participants indicated that introducing new Alien attributes and dimensions would make the game more interesting.

“Maybe you could add…. like a small spikey teeth, or one reading a book…make them do signs with their hands, like a thumbs up.” P12

“[You could have more attributes]..Like a tongue sticking out…some could be sad or crying…” P9

4. DISCUSSION

This section will first provide a summary of the results before outlining the clinical implications of the findings. It will then reflect on the strengths and limitations of the current study and how the reliability and validity could be developed in future testing.

4.1. Revisiting the Research Questions

This study aimed to update the Alien Game and provide a practical scoring system, built on ideas from Pavitt's (2017) study. The researcher aimed to address the following questions:

- Is the updated Aliens Game a useful measure of concept formation in children aged 8-11 years?
- Do typically developing children approach the test in a predictable, and exhibit range of scores that do have 'floor' or 'ceiling' effects?
- Do demographic characteristics, existing measures of concept formation, or behaviour ratings of participants reflect their performance on the new game format?

4.2. Summary of Findings

4.2.1. Is the updated Aliens Game a useful measure of concept formation in children aged 8-11 years?

The improvements to the overall Alien Game suggest that the game has the 'true' properties to assess concept formation skills. This was suggestive of participants' responses in which each attribute was asked at least once. The TQ and IQ game measures also showed a significant association with the non-verbal abstraction measure. Additionally, introducing a game script ensured a consistent and coherent delivery of game instructions.

The findings indicated that children in this sample understood the game instructions and drew upon similar search strategies while playing the Alien Game. The majority of the children asked a 'constraint 'seeking' question as early as the first question of

Trial 1, and this game strategy remained consistent over the trials. Within every trial, children's first question removed one-third of the aliens (i.e., 8 out of 24), which is indicative of effective concept formation and application.

It was also found that children who guessed the target Alien with the fewest number of questions also made the fewest number of errors (i.e., asked the fewest ineffective or unallowed questions). This confirms that children with good conceptual skills consistently asked 'constraint-seeking' questions. Whereas those with weaker concept formation asked many more questions to reach the target, and also asked more of ineffective questions. However, it is worth noting that even when they asked more questions to reach the target, the majority of their questions were 'constraint seeking'. Qualitatively, it appeared that when children asked more than, for example, 10 questions to identify the target Alien, this might have been because they 'restarted' the trial. When children asked ineffective questions (e.g., a repeated question) they often noticed their mistake (e.g., the child said, 'Oh no, I asked that before'). On the whole, the children made considered choices and took time to frame their subsequent questions, based on feedback. This shows that children utilised a similar and effective questioning strategy.

4.2.2. Do typically developing children approach the test in a predictable, exhibit range of scores do not have 'floor' or 'ceiling' effects?

Results showed that calculating the measures using absolute values produced the best range of scores, with no children obtaining particularly low or high scores, particularly for total questions. This indicates that the data collected provided a good distribution of scores, and that the game was not too easy or too hard for this group of children. So, Total Questions could be an effective way of capturing game performance. The absence of 'floor' and 'ceiling' effects indicates that it would be possible to develop the game further so psychometric descriptors (Urban, 1908), such as 'low average', 'above average', could be applied to capture aspects of concept formation skills.

4.2.3. Do demographic characteristics, and existing measures of concept formation, or behaviour ratings of participants reflect their performance on the new game format?

The research question was concerned with testing the concurrent validity and predictive of the Alien Game. The researcher aimed to investigate whether performance on Alien Game correlated with measures of verbal reasoning (i.e., Similarities) and visual reasoning (i.e., Matrix Reasoning skills), and the teacher-rated measure of executive functions in everyday behaviour [predictive validity]; (i.e., CHEXI). Sex and language of participants were used as separate variables to see whether any of these independent variables were associated with the test scores.

The results showed that Alien Game measures were not correlated with verbal abstraction skills. However, the TQ, IQ and WI measures were correlated with Matrix reasoning, suggesting that the Alien Game measures non-verbal abstraction.

The CHEXI Working Memory and Inhibition domains were not associated with any of the Alien Game Measures. Only small associations were found between Working Memory and IQ's, WI and AG game measures, and between Inhibition and TQ and Ab measures. Overall, it appears that the teacher-report measures of executive functions and the Alien game are unrelated and measure different constructs. Perhaps the use of CHEXI was not the best measure of behaviours associated with executive functions, as certain items (e.g., 'Has difficulty refraining from smiling or laughing in situations where it is inappropriate') capture more of the emotional and social than the cognitive aspects of executive functions (Thorell & Catale, 2014).

4.2.3.1. Sex and Language Groups

Although there were no substantive differences between males and females in the Alien Game, female participants scored slightly better on all measures. For instance, females asked fewer TQ's and IQ's, obtained higher Ab scores, and consequently achieved higher scores on the WA, WI, and AG game measures. Females reached the target Alien in fewer questions, and their overall game score was higher than males. It is unclear why females obtained better results.

There were no associations between primary language and Alien Game measures. However, the EPL group scored slightly higher on all game measures apart from Ab. It could be that children who speak English as a primary language found it easier to articulate their thoughts and thus asked more effective questions. This seems to be in line with studies that suggest neuropsychological tasks tend to favour primary English speakers (Kisser et al., 2012). Despite this, no significant differences between the EPL and EAL groups in this study, suggested that the Alien Game is equally accessible for children who speak English as an additional language.

4.3. Comparisons to Previous Literature

Pavitt (2017) used proportions to calculate game scores. Both proportions and absolute values for calculated for the present study, and it was found that absolute values were a better way to calculate the measures, due to the simplicity of scoring and the greater range in scores. Similar to Pavitt's (2017) study the abstraction score consisted of the first three questions asked across trials 1-4 in their scoring. This is because the children in this sample also managed to complete a trial with only three questions. Unlike the rules used in Pavitt's (2017) and Alderson-Day and McGonigle-Chalmers (2011) study, a 10-question cut-off point was not applied for this sample of children since the research is still in the early stages of test development, and concerned with capturing data with the potential to develop norms in future. A 10-question limit could be adopted in the final build of the game, if necessary.

Consistent with results from the 20-Questions task and procedures, used in Alderson-Day and McGonigle-Chalmers study, the children in this study also asked predominantly 'constraint seeking' questions. The current research revised Pavitt's game scoring codes and introduced new codes. All questions other than 'constraint seeking' were coded as ineffective questions (this code included 'hypothesis seeking' questions, also). This single code captured all 'unallowed' or ineffective questions asked within a trial. Similar to the D-KEFS scoring system, a weighted score for ineffective questions were created, similar to the WA score, which was not included in Pavitt's scoring system. Additionally, an overall game performance code

was created, comprising weighted total questions and weighted ineffective questions scores.

The current study produced a low resource layout of the game, which meant that the game was more accessible to children than presenting measures on dedicated apparatus (e.g., Condy et al., 2021). Although the children in this study had no motor difficulties, in future studies the layout of the game should be accessible to children with physical impairments. Unlike Condy et al., and Alderson-May and McGongle-Chalmer's study, this present study only tested typically developing children. The results would have differed if the sample included a group of children with neurodevelopmental difficulties.

4.4. Critical Evaluation

4.4.1. Strengths of the Current Study

Design – this current study aimed to develop the images used in Pavitt (2017) and make them more appealing to young children, by refining the Alien attributes and making them more prominent. Content analysis showed that all 13 attributes were named by the children, meaning that the research was successful in refining the game and making each attribute salient. The current study also obtained qualitative feedback from participants on how the Alien Game could be developed; the suggestions could be applied in future versions of the Alien Game, to improve game enjoyment and performance. All children rated the Alien Game as being highly enjoyable. Half of the participants who provided feedback about how the game could be improved, stated that the game did not require any changes.

This study also recruited a diverse sample of children, with more than half of the children speaking English as an additional language; there was also an equal mix of sexes.

In order to reduce the language demands of the Alien Game, Alien names were removed, and instead assigned numbers for the purpose of making identification of the target Alien clear to the examiner. The numbers were positioned on the top left corner of each alien and were small so that children did not use the numbers as part of their game strategy. Indeed, assigning numbers proved to be useful; the majority

of the children referred to the number of the Alien when making a guess attempt (i.e., is it Alien number 4). This helped the examiner to capture the correct target Alien from each child and avoided any misunderstanding.

Scoring system - this current research also developed the scoring system used in Pavitt's (2017) study and successfully identified new ways to measure game performance. Firstly, the researcher calculated the measures in several ways using proportions and then absolute numbers. This enabled the researcher to see the strengths and weaknesses of both methods. All final measures included in the analysis comprised absolute values rather than proportions. Using absolute values increases the user friendliness of the test and makes the scoring system more accessible to professionals. In addition, TQ appears to be a good way to evaluate game performance. Again, this would be an easier measure to comprehend and calculate by most professionals without services having to provide extensive and expensive materials or training to staff.

Additionally, having an AG score means that a single score could capture overall game performance. This is similar to the scoring found in existing neuropsychological where a single score encapsulates scores from subdomains. Developing the WA score and introducing a weighted score for the IQ's is also a strength of this study as these measures enable individuals to score proportionally.

A further strength of the game materials is that there were no differences between sex and language groups for the Alien Game, nor for the established measures of verbal and non-verbal abstraction, and behaviours of executive functions. This suggests that the Alien Game is accessible for children from all backgrounds and genders.

4.4.2. Limitations

Regardless of the strengths of the present study, there were several limitations to note. Firstly, the sample came from a small one-form entry primary school and with a student-teacher ratio higher than the typical for mainstream schools in the UK. This is significant given that teacher input positively influences pupils' educational

progress, attainment, and behaviour. Children in this sample were verbally able group of children, which was evident from their verbal abstraction scores. None of the children had any verbal or visual deficits. These factors reduce the generalisability of the findings of this study as children from more deprived schools with lower verbal skills may perform differently, and thus change the findings about the feasibility of the Alien Game measure.

Although refining the images of the Aliens was important, one limitation here is that all target aliens were presented in the same order and on the same grid to each child. Therefore, it was difficult to establish whether any of the Aliens were easier or harder to identify due to the attributes assigned to them. It could be that some attributes made it harder to spot the target alien and resulted in the child asking more questions to reach the target. Although the numbers assigned to the Aliens to make identifying the target Alien easier, rather than it being used as a game strategy, this did permit one child using the numbers as their main game strategy. They also seemed to ask mostly spatial questions (e.g., is it on the left side of the game).

Another limitation of the Alien Game is that although the researchers aimed to even out the attributes, some aliens had an uneven number of dimensions assigned to them. Perhaps this was due to the researcher not using a more intricate and methodical strategy to even out and distribute the attributes. For instance, the tail attribute only had two dimensions (i.e., tail, no tails), while the nose attribute had three dimensions (i.e., no nose, small nose, big nose). This was of relevance to the target Alien's, as the target Alien in the fourth trial (i.e., Alien no. 24) had a greater number of dimensions compared to the target Alien in the second trial (i.e., Alien no. 8), which may have created a variation on the number of questions needed to be asked to reach the target. Future researchers could overcome this limitation by using a more systematic method to even out the attributes so that game performance is not inadvertently biased by the numbers of dimensions assigned to each Alien. Each game trial could consist of certain number of dimensions to evaluate the impact on game performance.

Although the significant correlation between the Alien Game (i.e., the TQ and IQ scores) and non-verbal abstraction measures are promising, it is important to

emphasise that this produced a small to medium association. Thus, higher correlation coefficients for concurrent validity should be aimed for in future versions of the game (e.g., 0.5 or above to be relatively strong). Further, it is worth noting that since the Alien Game shows no significant correlation with verbal abstraction, the game may be unable to identify children with verbal difficulties. Consequently, professionals may be limited in clinically capturing children with verbal deficits.

Finally, the researcher noticed that the children who completed the game in the afternoon took longer to complete the task than the sample of children who participated earlier in the day. In hindsight, it would have been useful to make a note of the timings of testing so group comparisons could be made regarding whether time of day impacts game performance due to changes in attention and motivation.

4.5. Clinical Implications

The second stage of this pilot study shows that the sample of children in this study understood the instructions of the test and used similar strategies on the Alien Game. This is promising as it indicates that normative characteristics can be established from using this test, and the game could be further developed to devise it into a formalised neuropsychological test. Given that the TQ measure correlated with non-verbal reasoning, it would not disadvantage children with developmental language delays or children who speak English as an additional language (Durant et al., 2019).

The NHS and local authorities are impacted by the austerity and services are often slow at identifying and effectively treating executive function difficulties in children (Cummins, 2019). Instead, behaviours related to weak executive functions are often attributed to poor behaviour or lack of interest in learning, and these children are disproportionately disadvantaged throughout their developmental trajectory without the relevant strategies put in place (Frazier et al., 2022). Therefore, having accessible test materials that are cheap to produce and require minimal staff training could enable professionals to identify needs and offer interventions in a timely manner. Additionally, assessing children in their natural environment (e.g., in schools) would reduce the burden placed upon parents, to attend services and

minimise the stigma associated with being assessed in clinical settings. This could reduce the health, educational and economic inequity for children from disadvantaged backgrounds in line with the NHS Long-Term Plan (NHS England, 2019).

4.6. Future Research Directions

4.6.1. Reliability

Given that the children in this study were more verbally able than children of similar age group, future researchers should aim to recruit a bigger sample of children with lower verbal ability. The likelihood of recruiting a more diverse ability of children could be enhanced by specifically avoiding recruitment of participants from one-form entry and private schools, as these schools do not reflect the education system accessed by most children since only a small proportion of children are able to access such schooling.

All participants completed four trials of the Alien Game. The learning slopes showed no substantive improvements or declines between trials two and four, suggesting a plateau had been achieved, and therefore minimising learning effects ($p>0.05$). It might be that administering one trial is enough to see concept formation difficulties, similar to the Brixton Test, which is only administered once. In order to test this empirical question, future researchers could introduce 5, 10 or 15 trials of the game, to see whether more trials lead to more consistent changes in game performance. This will help establish the optimum number of trials to administer. Additionally, it may be useful to conduct a test re-test reliability to see if the same group of children would benefit from exposure to the same test in an enduring way. Though the research would expect children to show some improvements, a future study could establish whether the benefit of time is greater than the benefit you would expect from practice by doing a reliability chain analysis.

To determine whether the Alien attributes impacts on game performance, future researchers could change the way the game is presented. For instance, the target Aliens could be presented in various sequences; and each trial could be presented

on a new grid, which would encourage children to scan the entire grid each time. Doing this would provide useful information about whether any of the Alien configurations were easier or harder to spot across the trials. In a future study, Alien attributes could be introduced as an influencing variable in the analysis. Time of testing (e.g., morning versus afternoon testing) could also be added as an influencing variable, to investigate the impact of time on test performance, though there will, undoubtedly, be individual differences. Analysis on the type of attributes asked showed that the most frequently asked question was the colour of the Aliens. Feedback from children indicated that asking about the colour of the Aliens eliminated a greater number of items. With this in mind, the game script could be developed to refrain participants from asking about the colour of the Alien as their first question.

A methodological limitation of the study is that the researchers involved in the development of the cartoon images of the Aliens could have biased the selection of attributes based on their own subjective judgments of what may be adequate for the game, which may be culturally influenced. Even though results showed that each attribute was indeed distinguishable, as each attribute was named at least once, cartoon images could be reviewed by a small sample of children in the future, to identify potential issues, before finalising the images. Additionally, the involvement of experts by experience (e.g., children accessing child mental health services) could be involved in the design of the game.

The game instruction could be refined to make it clear that the children could point on the grid if they do not know the name of the attributes, and that the Alien numbers should only be used to identify the target Alien. Examiners should have an image of the target Alien on the corresponding response sheets so that children's questions are not influenced by where the examiner is looking on the game grid, which may minimise a 'lucky guess'.

Children were aware that the researcher had devised the Alien Game, and this may have led them to respond more positively. In future, an electronic feedback form could be created for game feedback; this would not only enhance respondent anonymity, but also encourage more honest feedback.

4.6.2. Validity

The study aimed to test the feasibility and concurrent validity of the Alien Game measure by comparing the game scores against the established measures of concept formations (i.e., the Similarities and Matrix Reasoning in the WISC-IV subtests). These measures were chosen to see whether any of the children had verbal and visuo-spatial concept formation deficits.

An important consideration was to minimise any practice effects by comparing the performance of the multi-trial Alien Game with a single-trial test of concept formation, and the Similarities and Matrix Reasoning tests enabled this comparison as they are both single-trial measures of concept formation. A limitation of using these WISC-IV subtests is that they are both measures of abstraction, and it would have been useful to have chosen measures of induction. However, a difficulty in choosing a measure of induction is that the researcher is currently not aware of suitable measures of induction for children, and this limits the selection of such measures.

A measure of executive function (i.e., CHEXI) was used to make real world comparisons. Perhaps this was not the best way to capture behaviours associated with the executive function, especially when considering that boys are usually rated much worse on such measures (Timimi, 2017), leading to the 'pathologisation' of developmentally appropriate behaviours. It may be that future researchers would need to use a multi-trial test, such as the WCST, in order to address convergent validity, though this is a test that has been developed for both adults and children. The scores on the WCST could be compared against the game measures (e.g., TQ, AS, WA, WI, or AG) to check for correlations between the measures.

Although this study recruited a more diverse sample of children, a larger proportion of children from non-Western cultures could be recruited in future.

4.7. Research Reflexivity

The critical realist position adopted in this study aligns with my worldview, such that I believe there is a universal truth of 'executive' functions' even though this might be conceptualised differently across cultures. The differences in the way we conceptualise executive functions do not change the 'truth' that all humans are influenced by the same neural architectural foundation (Kelkar et al., 2013). Having said this, I do not believe that these neural pathways are 'fixed' and that they can be strengthened through bespoke interventions. This is particularly pertinent during childhood as children do not reach cognitive maturation until early adulthood (Ganesan & Steinbeis, 2022).

Indeed, the children in this study had good verbal and non-verbal skills and came from a range of backgrounds. I was struck by how skilled they were at completing the Similarities and Matrix Reasoning tasks, and quick at grasping the game instructions and utilising an effective game strategy. Perhaps one reason why children may have done so well in all test measures is because of the availability in the range of toys and games that are available now, which enable them to solve similar novel problems, and so the Similarities and Matrix Reasoning tasks are not as novel to children as they once were (Chierchia et al., 2019). Furthermore, this was a highly resourced school with a good student-teacher ratio, with smaller class sizes, meaning that children accessed good quality education. Though 'executive functions' skills are likely to vary from individual to individual, results from this study highlight the significant role of having access to resources and high-quality teaching and learning in promoting children's concept formation and problem solving skills, regardless of their background. In some ways, performance on cognitive tests is therefore impacted by environmental factors, which are also important aspects of good education and academic achievement (Koc & Celik, 2015). I feel these are important considerations when interpreting the results presented in this study.

4.8. Conclusions

The results from the current study have demonstrated that the updated Alien Game shows potential as a measure of concept formation for typically developing young

children aged between 8 and 11 years. In its current stage, the Alien Game is a valid measure of non-verbal receptive abstraction skills for both sexes and diverse language groups. Future research should continue to trial the Alien Game, potentially with a greater number of game trials to determine the optimum number of trials, recruiting children with lower verbal ability, and for the development of normative data in the future. The final build of the game could incorporate Total Questions as a measure of game performance, which would be easy to calculate by most clinicians working in fast-paced NHS environments. Continued research and development of the Alien Game could provide clinicians with measures of concept formation that are suitable for children with a range of language abilities and backgrounds, and offer benefits for children in receiving developmentally appropriate assessments and interventions.

5. REFERENCES

Adobe Inc. (2019) Adobe *Illustrator*. Retrieved from https://adobe.com/products/illustrator

Amukune, S., & Józsa, K. (2021). The Childhood Executive Functioning Inventory (CHEXI): Psychometric Properties and Association with Academic Achievement in Kenyan First Graders. *Journal of Psychological and Educational Research*, *29*(1), 154-176

Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological Rehabilitation*, *8*(3), 319–349.

Arguedas M, Xhafa F, Casillas L, Daradoumis T, Peña A, Caballé S (2018) A model for providing emotion awareness and feedback using fuzzy logic in online learning. Soft Comput *22*(3):963–977

Baldo, J. V., Delis, D. C., Wilkins, D. P., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Archives of Clinical Neuropsychology*, *19*(3), 407–419. Scopus. https://doi.org/10.1016/S0887-6177(03)00074-X

Barkan, J. H., & Bernal, E. M. (1991). Gifted Education for Bilingual and Limited English Proficient Students. *Gifted Child Quarterly*, *35*(3), 144–147. https://doi.org/10.1177/001698629103500306

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 10.1080/13854046.2012.663001. https://doi.org/10.1080/13854046.2012.663001

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function: Development of executive functions. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.0149

Bhaskar, R. (1990). Reclaiming Reality. Verso.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and*

Psychopathology, *20*(3), 821–843. https://doi.org/10.1017/S0954579408000394

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. Annual Review of Psychology, 66, 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Bodrova, E., & Leong, D. (1996). Tools of the Mind: The Vygotskian Approach to Early Childhood Education. Merrill.

Boehm, A. E. (2004). Assessment of basic relational concepts. In The psychoeducational assessment of preschool children, 3rd edition (pp. 186–203). Lawrence Erlbaum Associates Publishers.

Booth, A. E., & Waxman, S. R. (2002). Word learning is “smart”: Evidence that conceptual information affects preschoolers’ extension of novel words. *Cognition*, *84*(1), B11-22. https://doi.org/10.1016/s0010-0277(02)00015-x

Bott, L., Brock, J., Brockdorff, N., Boucher, J., & Lamberts, K. (2006). Perceptual similarity in autism. *The Quarterly Journal of Experimental Psychology*, *59*, 1237–1254. https://doi.org/10.1080/02724980543000196

Bracken, B. A., & Cato, L. A. (1986). Rate of conceptual development among deaf preschool and primary children as compared to a matched group of nonhearing impaired children. *Psychology in the Schools*, 23, 95–99. https://doi.org/10.1002/1520-6807(198601)23:1<95:AID-PITS2310230115>3.0.CO;2-L

Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children’s Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, *19*(3), 273–293. ttps://doi.org/10.1207/S15326942DN1903_3

Bynner, J. M. (1997). Basic skills in adolescents’ occupational preparation. *The Career Development Quarterly*, 45, 305–321. https://doi.org/10.1002/j.2161-0045.1997.tb00536.x

Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F., & Segall, M. J. (2008). Evidence-based assessment of cognitive functioning in paediatric psychology. *Journal of Paediatric Psychology*, *33*(9): 999–1014. https//doi.org/10.1093/jpepsy/jsm138

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, 27, 270-295. https://doi.org/10.1006/ceps.2001.1094

Castellanos, I., Kronenberger, W. G., Beer, J., Colson, B. G., Henning, S. C., Ditmars, A., & Pisoni, D. B. (2015). Concept Formation Skills in Long-Term Cochlear Implant Users. J*ournal of Deaf Studies & Deaf Education*, *20*(1), 27–40. CINAHL Complete.

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

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. J*ournal of Attention Disorders*, *19*(6), 489–495. https://doi.org/10.1177/1087054712470971

Cerrato, A., & Ponticorvo, M. (2017). Enhancing Neuropsychological Testing with Gamification and Tangible Interfaces: The Baking Tray Task. In J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo, & H. Adeli (Eds.), *Biomedical Applications Based on Natural and Artificial Computing* (Vol. 10338, pp. 147–156). Springer International Publishing. https://doi.org/10.1007/978-3-319-59773-7_16

Cheng, V. W. S., Davenport, T., Johnson, D., Vella, K., & Hickie, I. B. (2019). Gamification in Apps and Technologies for Improving Mental Health and Well-Being: Systematic Review. *JMIR Mental Health*, *6*(6), e13717. https://doi.org/10.2196/13717

Chevignard, M. P., Catroppa, C., Galvin, J., & Anderson, V. (2010). Development and Evaluation of an Ecological Task to Assess Executive Functioning Post Childhood TBI: The Children's Cooking Task. *Brain Impairment*, *11*(2), 125–143. https://doi.org/10.1375/brim.11.2.125

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, *6*(10), 190232. https://doi.org/10.1098/rsos.190232

Cirino, P. T., Ahmed, Y., Miciak, J., Taylor, W. P., Gerst, E. H., & Barnes, M. A. (2018). A framework for executive function in the late elementary years.

Neuropsychology, *32*(2), 176–189. APA PsycInfo. https://doi.org/10.1037/neu0000427

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, *46*(5), 1176–1191. https://doi.org/10.1037/a0019672

Clark, A. E. (1997). Job Satisfaction and Gender: Why Are Women So Happy at Work? *Labour Economics*, 4, 341-372. http://dx.doi.org/10.1016/S0927-5371(97)00010-9

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*, 103993. https://doi.org/10.1016/j.ridd.2021.103993

Cummins, I. (2018). The Impact of Austerity on Mental Health Service Provision: A UK Perspective. *International Journal of Environmental Research and Public Health*, *15*(6), 1145. https://doi.org/10.3390/ijerph15061145

Dekkers, T. J., Agelink van Rentergem, J. A., Koole, A., van den Wildenberg, W. P. M., Popma, A., Bexkens, A., Stoffelsen, R., Diekmann, A., & Huizenga, H. M. (2017). Time-on-task effects in children with and without ADHD: Depletion of executive resources or depletion of motivation? *European Child & Adolescent Psychiatry*, *26*(12), 1471–1481. https://doi.org/10.1007/s00787-017-1006-y

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis–Kaplan Executive Function System [D-KEFS]*. San Antonio: The Psychological Corporation, Harcourt Brace & Company

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011 (Vol. 11, p. 15). https://doi.org/10.1145/2181037.2181040

Diamond, A. (2020). Chapter 19—Executive functions. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 225–240). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Diamond, A. (2020a). Chapter 19—Executive functions. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), Handbook of Clinical Neurology (Vol. 173, pp. 225–240). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Diamond, A. (2020b). Chapter 19—Executive functions. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), Handbook of Clinical Neurology (Vol. 173, pp. 225–240). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

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

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 Psycholog*y, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Durant, K., Peña, E., Peña, A., Bedore, L. M., & Muñoz, M. R. (2019). Not All Nonverbal Tasks Are Equally Nonverbal: Comparing Two Tasks in Bilingual Kindergartners With and Without Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3462–3469. https://doi.org/10.1044/2019_JSLHR-L-18-0331

Erikson, E.H. (1963) Chapter 6. In *Childhood and Society*. New York: Norton.

Floyd, R. G., McCormack, A. C., Ingram, E. L., Davis, A. E., Bergeron, R., & Hamilton, G. (2006). Relations between the Woodcock-Johnson III clinical clusters and measures of executive functions from the Delis-Kaplan executive function system. *Journal of Psychoeducational Assessment*, *24*, 303-317. https//doi.org/10.1177/0734282906287823

Frazier, T. W., Crowley, E., Shih, A., Vasudevan, V., Karpur, A., Uljarevic, M., & Cai, R. Y. (2022). Associations between executive functioning, challenging behavior, and quality of life in children and adolescents with and without neurodevelopmental conditions. *Frontiers in Psychology*, *13*, 1022700. https://doi.org/10.3389/fpsyg.2022.1022700

Freud, A. (1968). The Ego and the Mechanisms of Defence. International Universities Press.

Ganesan, K., & Steinbeis, N. (2022). Development and plasticity of executive functions: A value-based account. *Current Opinion in Psychology*, 44, 215–219. https://doi.org/10.1016/j.copsyc.2021.09.012

Gelman, S. A. (2009). Learning from Others: Children's Construction of Concepts. *Annual Review of Psychology*, *60*(1), 115–140. https://doi.org/10.1146/annurev.psych.59.103006.093659

Gerring, J. P., & Wade, S. (2012). The Essential Role of Psychosocial Risk and Protective Factors in Pediatric Traumatic Brain Injury Research. *Journal of Neurotrauma*, *29*(4), 621–628. https://doi.org/10.1089/neu.2011.2234

Gioia, G. A. (2015) Multimodal evaluation and management of children with concussion: Using our heads and available evidence. *Brain Injury*, *29*(2), 195-206. https//doi.org/10.3109/02699052.2014.965210

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238. https//doi.org/10.1076/chin.6.3.235.3152

Gligorović, M., & Buha, N. (2013). Conceptual abilities of children with mild intellectual disability: Analysis of Wisconsin Card Sorting Test performance. *Journal of Intellectual & Developmental Disability*, *38*(2), 134–140. https://doi.org/10.3109/13668250.2013.772956

Glutting, J. J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test manual*, Wilmington, DE: Wide Range.

Goldstein, F. C., & Levin, H. S. (1991). Question-asking strategies after severe closed head injury. *Brain and Cognition*, *17*(1), 23–30. https://doi.org/10.1016/0278-2626(91)90063-e

Gómez-Tello, M. F., Rosetti, M. F., Galicia-Alvarado, M., Maya, C., & Apiquian, R. (2022). Neuropsychological screening with TOWI: Performance in 6- to 12-year-old children. *Applied Neuropsychology: Child*, *11*(2), 115–124. https://doi.org/10.1080/21622965.2020.1764357

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications*, *7*(1), 11. https://doi.org/10.1186/s41235-021-00350-w

Gorham, D. R. (1956). A proverbs test for clinical and experimental use. *Psychological Reports*, *2*(1), 1-12. https//doi.org/10.2466/PR0.2..1-12

Greenwood, J. D. (1994). *Realism, identity and emotion: Reclaiming social psychology*. Sage.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development*, *70*(4), 817–832. https://doi.org/10.1111/1467-8624.00059

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scafiddi, N. T., & Tonks, S. (2004). Increasing Reading Comprehension and Engagement Through Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, *96*(3), 403.

Hamilton, C. J., Coates, R. O., & Heffernan, T. (2003). What develops in visuo-spatial working memory development? *European Journal of Cognitive Psychology*, 15, 43–69. https://doi.org/10.1080/09541440303597

Hammill, D. D., & Pearson, N. A. (2009). Comprehensive Test of Nonverbal Intelligence—Second Edition. I*n Practitioner's guide to assessing intelligence and achievement* (pp. 233–264). John Wiley & Sons Inc.

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, *14*(1), 91–99.

Heaton, R., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test (WCST) manual revised and expanded*. Odessa, FL: Psychological Assessment Resources. https//10.1037/t31298-000

Hessl, D., Sansone, S. M., Berry-Kravis, E., Riley, K., Widaman, K. F., Abbeduto, L., Schneider, A., Coleman, J., Oaklander, D., Rhodes, K. C., & Gershon, R. C. (2016). The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. *Journal of Neurodevelopmental Disorders*, *8*(1), 35. https://doi.org/10.1186/s11689-016-9167-4

Hick, R., Botting, N., & Conti-Ramsden, G. (2005). Cognitive abilities in children with specific language impairment: Consideration of visuo-spatial skills. *International Journal of Language & Communication Disorders*, 40, 137–149. https://doi.org/10.1080/13682820400011507

Hill, E. L. (2004). Executive dysfunction in autism☆. *Trends in Cognitive Sciences*, *8*(1), 26–32. https://doi.org/10.1016/j.tics.2003.11.003

Hoffman, L. M., & Gillam, R. B. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech,*

Language, and Hearing Research, *47*(1), 114–125. https://doi.org/10.1044/1092-4388(2004/011)

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology*, 11. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00721

Howard, S. J., Okely, A. D., & Ellis, Y. G. (2015). Evaluation of a differentiation model of preschoolers' executive functions. *Frontiers in Psychology*, 6, 285. https://doi.org/10.3389/fpsyg.2015.00285

Huffman, L. C., Mehlinger, S. L., Kerivan, A. S., Cavanaugh, D. A., Lippitt, J., & Moyo, O. (2001). *Off to a Good Start: Research on the Risk Factors for Early School Problems and Selected Federal Policies Affecting Children's Social and Emotional Development and Their Readiness for School*. Eric. https://eric.ed.gov/?id=ED476378

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Huizinga, Johan. 1955. *Homo Ludens: A Study of the Play-element in Cult*. Boston: Beacon Press.

Hwang, J., Mitz, A. R., & Murray, E. A. (2019). NIMH MonkeyLogic: Behavioral control and data acquisition in MATLAB. *Journal of Neuroscience Methods*, 323, 13–21. https://doi.org/10.1016/j.jneumeth.2019.05.002

IBM Corp. Released 2021. *IBM SPSS Statistics for Windows, Version 28.0*. Armonk, NY: IBM Corp

Kagan, J. (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology, 71*(1), 17–24. https://doi.org/10.1037/h0022886

Kelkar, A. S., Hough, M. S., & Fang, X. (2013). Do we think alike? A cross-cultural study of executive functioning. *Culture and Brain*, *1*(2), 118–137. https://doi.org/10.1007/s40167-013-0010-4

Klausmeier, H. J., & Hooper, F. H. (1974). 1: Conceptual Development and Instruction. *Review of Research in Education*, *2*(1), 3–54. https://doi.org/10.3102/0091732X002001003

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *13*(1), 111-124. https//doi.org/10.1017/S0954579401001080

Koc, N., & Celik, B. (2015). The Impact of Number of Students per Teacher on Student Achievement. Procedia - *Social and Behavioral Sciences*, 177, 65–70. https://doi.org/10.1016/j.sbspro.2015.02.335

Kocadere, S. A., & Çağlar, Ş. (2015). The design and implementation of a gamified assessment. *Journal of E-Learning and Knowledge Society*, *11*(3). https://doi.org/10.20368/1971-8829/1070

Langer, J. H. (1967). Vocabulary and Concept Development. *Journal of Reading*, *10*(7(, 448-456.

Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., & Farah, M. J. (2018). Childhood socioeconomic status and executive function in childhood and beyond. *PLoS ONE*, *13*(8), e0202964. https://doi.org/10.1371/journal.pone.0202964

Lee, J. J. & Hammer, J. (2011). Gamification in Education: What, How, Why Bother? Academic Exchange Quarterly, 15(2).

Levy, L., Lambeth, A., Solomon, R., & Gandy, M. (2018). Method in the madness: The design of games as valid and reliable scientific tools. *Proceedings of the 13th International Conference on the Foundations of Digital Games*, 1–10. https://doi.org/10.1145/3235765.3235793

Luckner, J. L., & McNeill, J. H. (1994). Performance of a group of deaf and hard-of-hearing students and a comparison group of hearing students on a series of problem-solving tasks. *American Annals of the Deaf*, 139, 371–377. https://doi.org/10.1353/aad.2012.0290

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, *4*(2), e11. https://doi.org/10.2196/games.5888

Lunkenheimer, E., Panlilio, C., Lobo, F. M., Olson, S. L., & Hamby, C. M. (2019). Preschoolers' Self-Regulation in Context: Task Persistence Profiles with Mothers and Fathers and Later Attention Problems in Kindergarten. *Journal of Abnormal Child Psychology*, *47*(6), 947–960. https://doi.org/10.1007/s10802-019-00512-x

Marschark, M., & Everhart, V. S. (1999). Problem-solving by deaf and hearing students: Twenty questions. *Deafness & Education International*, *1*(2), 65–82. https://doi.org/10.1002/dei.48

Martínez, L., Prada, E., Satler, C., Tavares, M. C. H., & Tomaz, C. (2016). Executive Dysfunctions: The Role in Attention Deficit Hyperactivity and Post-traumatic Stress Neuropsychiatric Disorders. *Frontiers in Psychology*, 7, 1230. https://doi.org/10.3389/fpsyg.2016.01230

Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language & Communication Disorders*, *43*(2), 181–200. https://doi.org/10.1080/16066350701340719

McCarthy, J. M., & Goffin, R. D. (2005). Selection Test Anxiety: Exploring Tension and Fear of Failure Across the Sexes in Simulated Selection Scenarios. *International Journal of Selection and Assessment*, *13*(4), 282–295. https://doi.org/10.1111/j.1468-2389.2005.00325.x

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, *3*, 341-350. https/doi.org/10.1016/0010-0277(74)90003-1

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35, 113–138. https://doi.org/10.1146/annurev.ps.35.020184.000553

Micalizzi, L., Brick, L. A., Flom, M., Ganiban, J. M., & Saudino, K. J. (2019). Effects of socioeconomic status and executive function on school readiness across levels of household chaos. *Early Childhood Research Quarterly*, 47, 331–340. https://doi.org/10.1016/j.ecresq.2019.01.007

Minshew, N., Siegel, D., Goldstein, G., & Weldy, S. (1994). Verbal problem solving in high functioning autistic individuals. *Archives of Clinical Neuropsychology*, *9*(1), 31–40. https//doi.org/10.1016/0887-6177(94)90012-4

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex "'Frontal Lobe'" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA

Statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the tower of London test. *Neuropsychologia*, *31*(12), 1367–1378. https://doi.org/10.1016/0028-3932(93)90104-8

Morton, N. W., & Preston, A. R. (2021). Concept formation as a computational cognitive process. *Computational Cognitive Neuroscience*, 38, 83–89. https://doi.org/10.1016/j.cobeha.2020.12.005

Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. Olver, & P. Greenfield, (Eds.), *Studies in cognitive growth* (pp. 86–102). New York: Wiley.

Nelson, H. E. (1976). A Modified Card Sorting Test Sensitive to Frontal Lobe Defects. *Cortex*, *12*(4), 313–324. https://doi.org/10.1016/S0010-9452(76)80035-4

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, *22*(4), 287-293. https//doi.org/10.1111/j.1745-3984.1985.tb01065.x

Nicolaidou, I., Tozzi, F., & Antoniades, A. (2022). A gamified app on emotion recognition and anger management for pre-school children. *International Journal of Child-Computer Interaction*, *31*, 100449. https://doi.org/10.1016/j.ijcci.2021.100449

NHS England. (2019). *The NHS Long Term Plan*. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf

Noble, K. G., & Giebler, M. A. (2020). The Neuroscience of Socioeconomic Inequality. *Current Opinion in Behavioral Sciences*, 36, 23–28. https://doi.org/10.1016/j.cobeha.2020.05.007

Osgood, C. E. (1953). *Method and Theory in Experimental Psycholog*y. Oxford University Press.

Palanivel, V., & Burrough, M. (2021). Acquired brain injury in children, and their rehabilitation: Where we are now? *Paediatrics and Child Health*, *31*(5), 176–180. https://doi.org/10.1016/j.paed.2021.02.001

Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken School Readiness Assessment for predicting first grade readiness. *Psychology in the Schools*, *46*(5), 397–409. https://doi.org/10.1002/pits.20385

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). Retrieved from http://uel.da.ulcc.ac.uk/6736/1/Anne Pavitt.pdf

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: Parent-report and cognitive assessment. *Developmental Medicine and Child Neurology*, *49*(6), 412–416. https://doi.org/10.1111/j.1469-8749.2007.00412.x

Pettifor, J. L. (1968). The role of language in the development of abstract thinking: A comparison of hard-of-hearing and normal-hearing children on levels of conceptual thinking. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, 22, 139–156. https://doi.org/10.1037/h0082755

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of research in science teaching*, *2*(3), 176-186.

Raban, D., & Geifman, D. (2009). Designing Online Information Aggregation and Prediction Markets for MBA Courses. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, 5, 247–262. https://doi.org/10.28945/76

Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive*. (Unpublished master's thesis). University of London, London.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, United Kingdom: Oxford Psychologists Press. https//doi.org/10.1037/t07027-000

Rivera-Batiz, F. L. (1992). Quantitative Literacy and the Likelihood of Employment among Young Adults in the United States. *The Journal of Human Resources*, *27*(2), 313–328. https://doi.org/10.2307/145737

Ropar, D., & Peebles, D. (2007). Sorting preference in children with autism: The dominance of concrete features. *Journal of Autism and Developmental Disorder*s, *37*(2), 270–280. https://doi.org/10.1007/s10803-006-0166-2

Satz, P. & Mogel, S. (1962). An abbreviation of the WAIS in clinical use. *Journal of Clinical Psychology*, 18, 77-79. https//doi.org/10.1002/1097-4679(196307)19:3<298::AID-JCLP2270190311>3.0.CO;2-G

Sergeant, J. A., Oosterlaan, J., & Van der Meere, J. J. (1999). Information processing and energetical factors in ADHD. In H. Quay & A. E. Hogan (Eds.), *Handbook of disruptive behavior* (pp. 75–104). Plenum Press.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology*, 70, 365–380. https://doi.org/10.1111/j.2044-8295.1979.tb01706.x

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry*, *57*(11), 1301–1309. https://doi.org/10.1016/j.biopsych.2005.01.043

Shu, C., Lung, W., Tien, A., & Chen, C. (2001). Executive function deficits in non-retarded autistic children. *Autism*, *5*(2), 165–174. https://doi.org/10.1177/1362361301005002006

Sokol, B. W., & Müller, U. (2007). The development of self-regulation: Toward the integration of cognition and emotion. *Cognitive Development*, *22*(4), 401–405. https://doi.org/10.1016/j.cogdev.2007.08.008

Song, H., Yi, D.-J., & Park, H.-J. (2020). Validation of a mobile game-based assessment of cognitive control among children and adolescents. *PLOS ONE*, *15*(3), e0230498. https://doi.org/10.1371/journal.pone.0230498

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, *63*(3–4), 289–298. https://doi.org/10.1007/s004269900007

Sukhodolsky, D. G., Smith, S. D., McCauley, S. A., Ibrahim, K., & Piasecka, J. B. (2016). Behavioral Interventions for Anger, Irritability, and Aggression in Children and Adolescents. *Journal of child and adolescent psychopharmacology*, *26*(1), 58–64. https://doi.org/10.1089/cap.2015.0120

Teubner-Rhodes, S., Vaden, K. I., Dubno, J. R., & Eckert, M. A. (2017). Cognitive persistence: Development and validation of a novel measure from the Wisconsin Card Sorting Test. *Neuropsychologia*, 102, 95–108. https://doi.org/10.1016/j.neuropsychologia.2017.05.027

Thorell, L. B., & Catale, C. (2014). The assessment of executive functioning using the Childhood Executive Functioning Inventory (CHEXI). In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 359–366). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-8106-5_20

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

Timimi, S., & Timimi, L. (2015). The social construction of attention deficit hyperactivity disorder. In M. O'Reilly & J. N. Lester (Eds.), The Palgrave handbook of child mental health (pp. 139-157). London: Palgrave Macmillan. https://doi.org/10.1057/9781137428318_8

Tröster, A. I. (2000). Clinical Neuropsychology, Functional Neurosurgery, and Restorative Neurology in the Next Millennium: Beyond Secondary Outcome Measures. *Brain and Cognition*, *42*(1), 117–119. https://doi.org/10.1006/brcg.1999.1178

Tulsky, D. S., & Heinemann, A. W. (2017). The clinical utility and construct validity of the NIH Toolbox Cognition Battery (NIHTB-CB) in individuals with disabilities. *Rehabilitation Psychology*, *62*(4), 409–412. https://doi.org/10.1037/rep0000201

Turan, Z., & Meral, E. (2018). Game-Based Versus to Non-Game-Based: The Impact of Student Response Systems on Students'Achievements, Engagements and Test Anxieties. *Informatics in Education*, 17. https://doi.org/10.15388/infedu.2018.07

Twomey, D. M., Wrigley, C., Ahearne, C., Murphy, R., De Haan, M., Marlow, N., & Murray, D. M. (2018). Feasibility of using touch screen technology for early cognitive assessment in children. *Archives of Disease in Childhood*, *103*(9), 853–858. https://doi.org/10.1136/archdischild-2017-314010

Urbanski, M., Bréchemier, M.-L., Garcin, B., Bendetowicz, D., Thiebaut de Schotten, M., Foulon, C., Rosso, C., Clarençon, F., Dupont, S., Pradat-Diehl, P., Labeyrie, M.-A., Levy, R., & Volle, E. (2016). Reasoning by analogy requires the left frontal pole: Lesion-deficit mapping and clinical implications. *Brain*, *139*(6), 1783–1799. https://doi.org/10.1093/brain/aww072

Vygotsky, L. S. (1986). *Thought and language*. (A. Kozulin Trans.), Cambridge, MA.: MIT Press. (Original work published 1934).

Wechsler, D. (2003). *The Wechsler intelligence scale for children [WISC-IV] (4th ed.)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D., Kaplan, E., Fein, D., Morris, R., Kramer, J. H., Maerlender, A., & Delis, D. C. (2004). *The Wechsler Intelligence Scale for Children-Fourth Edition Integrated*. San Antonio, TX: The Psychological Corporation

Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale [WISC-IV] (4th ed.)*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *The Wechsler intelligence scale for children [WISC-V] (5th ed.)*. San Antonio, TX: The Psychological Corporation.

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

Weeks, S. J., & Hobson, R. P. (1987). The salience of facial expression for autistic children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *28(*1), 137–151. https://doi.org/10.1111/j.1469-7610.1987.tb00658.x

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*(1), 43–53. https://doi.org/10.1037/a0016738

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*(3), 199–230. https://doi.org/10.1080/87565648809540405

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-old children. *Journal of Experimental Child Psychology*, *108*(3), 436–452. https://doi.org/10.1016/j.jecp.2010.08.008

Wieder, S. (2017). The Power of Symbolic Play in Emotional Development Through the DIR Lens*. Topics in Language Disorders, 37, 259–281.*

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biological Psychiatry*, *57*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006

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

Wilson, S., & Worsley, A. (2021). Unequal childhoods: A case study application of Lareau's 'accomplishment of natural growth' in British working-class and poor families. *British Educational Research Journal*, *47*(3), 770–786. https://doi.org/10.1002/berj.3707

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29. https://doi.org/10.1007/BF01531288

Wong, M. W. (2021). Fostering musical creativity of students with intellectual disabilities: Strategies, gamification and re-framing creativity. *Music Education Research*, *23*(1), 1–13. https://doi.org/10.1080/14613808.2020.1862777

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside

Wu L., Kim M., Markauskaite L. (2020). Developing young children's empathic perception through digitally mediated interpersonal experience: Principles for a hybrid design of empathy games. *British Journal of Educational Technology*, *51*(4),1168-1187.

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive Function: Implications for Education*. NCER 2017-2000. In National Center for Education Research. National Center for Education Research. https://eric.ed.gov/?id=ED570880

Zevenbergen, R., Hyde, M., & Power, D. (2001). Language, arithmetic word problems, and deaf students: Linguistic strategies used to solve tasks. Mathematics Education *Research Journa*l, *13*(3), 204–218. https://doi.org/10.1007/BF03217109

6. APPENDECIES

A Alien Game 6X4 Grid

B Visual Analogue Scale

C CHEXI

D Research Flyer

E Alien Pictures Used in Pavitt (2017)

F List of 13 Alien Attributes Used in Current Study

G Alien with Superordinate Attribute

H Thematic Mapping of Initial Codes

I Thematic Mapping of Final Codes

J Ethics Approval

K Organisation Information Sheet

L Organisation Loco Parentis Consent Form

M Parent Information Sheet

N Child Information Sheet

O Parent Consent Opt-Out Form

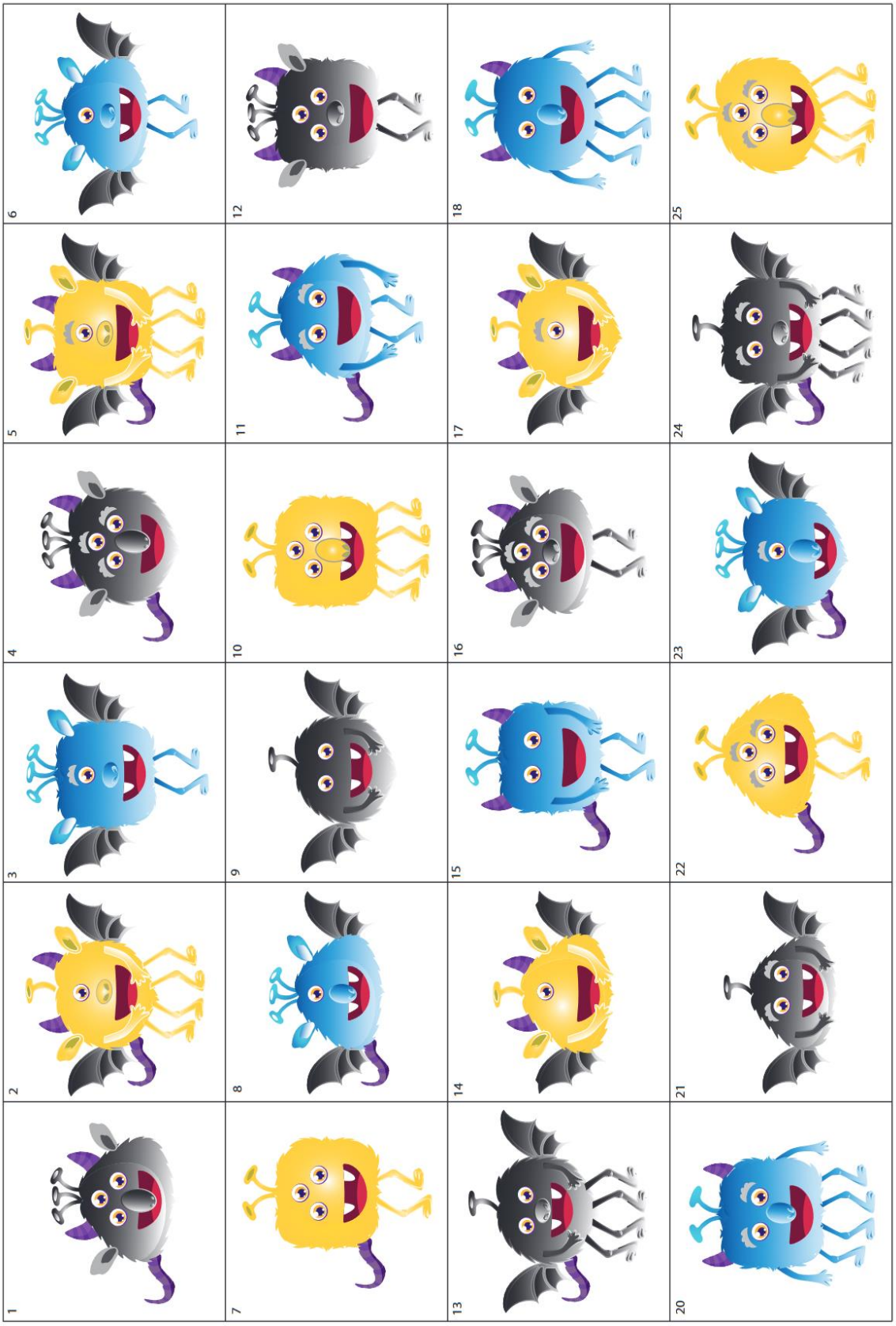
P Child Consent Form

Q Child Debrief Sheet

R Demographics Sheet

S Response Sheet

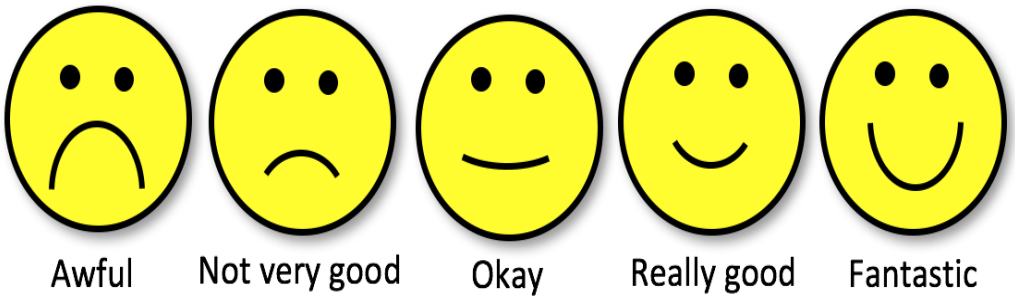
APPENDIX A: Alien Game 6X4 Grid



APPENDIX B: 5-point Likert Visual Analogue Scale

How enjoyable was the task today?

Tick the face that shows how you felt:



APPENDIX C: CHEXI

CHILDHOOD EXECUTIVE FUNCTIONING INVENTORY (CHEXI)

Please read each statement carefully and then say how true that statement if for the child. Indicate your response by circling one of the numbers (from 1 to 5) after each statement.

Child code or initials: Teacher:	1= Definitely not true. 2= Not true. 3= Partially true 4= True. 5= Definitely true	Definitely not true	Not true	Partially true	True	Definitely true
1. Has difficulty remembering lengthy instructions.		1	2	3	4	5
2. Seldom seems to be able to motivate themselves to do something that they don't want to do.		1	2	3	4	5
3. Has difficulty remembering what they are doing, in the middle of an activity.		1	2	3	4	5
4. Has difficulty following through on less appealing tasks unless they are promised some type of reward for doing so.		1	2	3	4	5
5. Tends to do things without first thinking about what could happen.		1	2	3	4	5
6. When asked to do several things, 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 they get stuck.		1	2	3	4	5
8. When something needs to be done, is often distracted by something more appealing.		1	2	3	4	5
9. Easily forgets what they have been 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 they find 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 their 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 their homework independently).		1	2	3	4	5
15. In order to be able to concentrate, they must find the task appealing.		1	2	3	4	5
16. Has difficulty refraining from smiling or laughing in situations where it is inappropriate.		1	2	3	4	5
17. Has difficulty telling a story about something that has happened so that others may easily understand.		1	2	3	4	5
18. Has difficulty stopping an activity immediately upon being told to do so (e.g., needs to jump a few extra times, or play on the computer a little bit longer after being asked to stop).		1	2	3	4	5
19. Has difficulty understanding verbal instructions unless they have also been shown how to do something.		1	2	3	4	5
20. Has difficulty with tasks or activities that involve several steps.		1	2	3	4	5
21. Has difficulty thinking ahead or learning from experience.		1	2	3	4	5
22. Acts in a wilder way compared to other children in a group (e.g., at a birthday party or during a group activity)		1	2	3	4	5
23. Has difficulty doing things that require mental effort, such as counting backwards.		1	2	3	4	5
24. Has difficulty keeping things in mind while they are doing something else.		1	2	3	4	5

APPENDIX D: Research Flyer

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

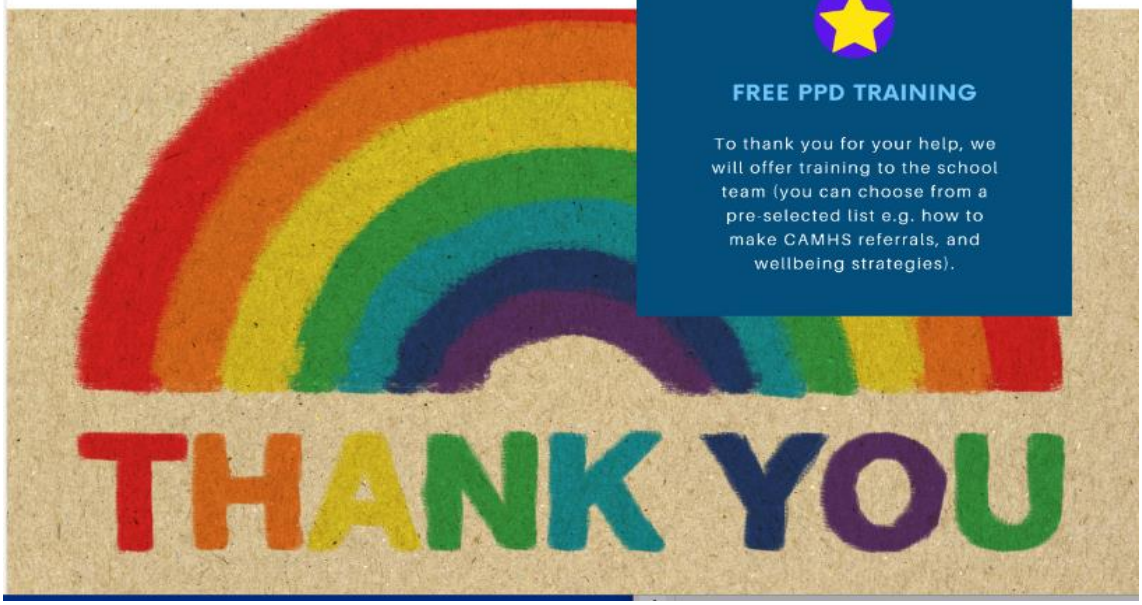
Pupils from Years 1 to Year 6 can participate with parent's consent.

STUDY

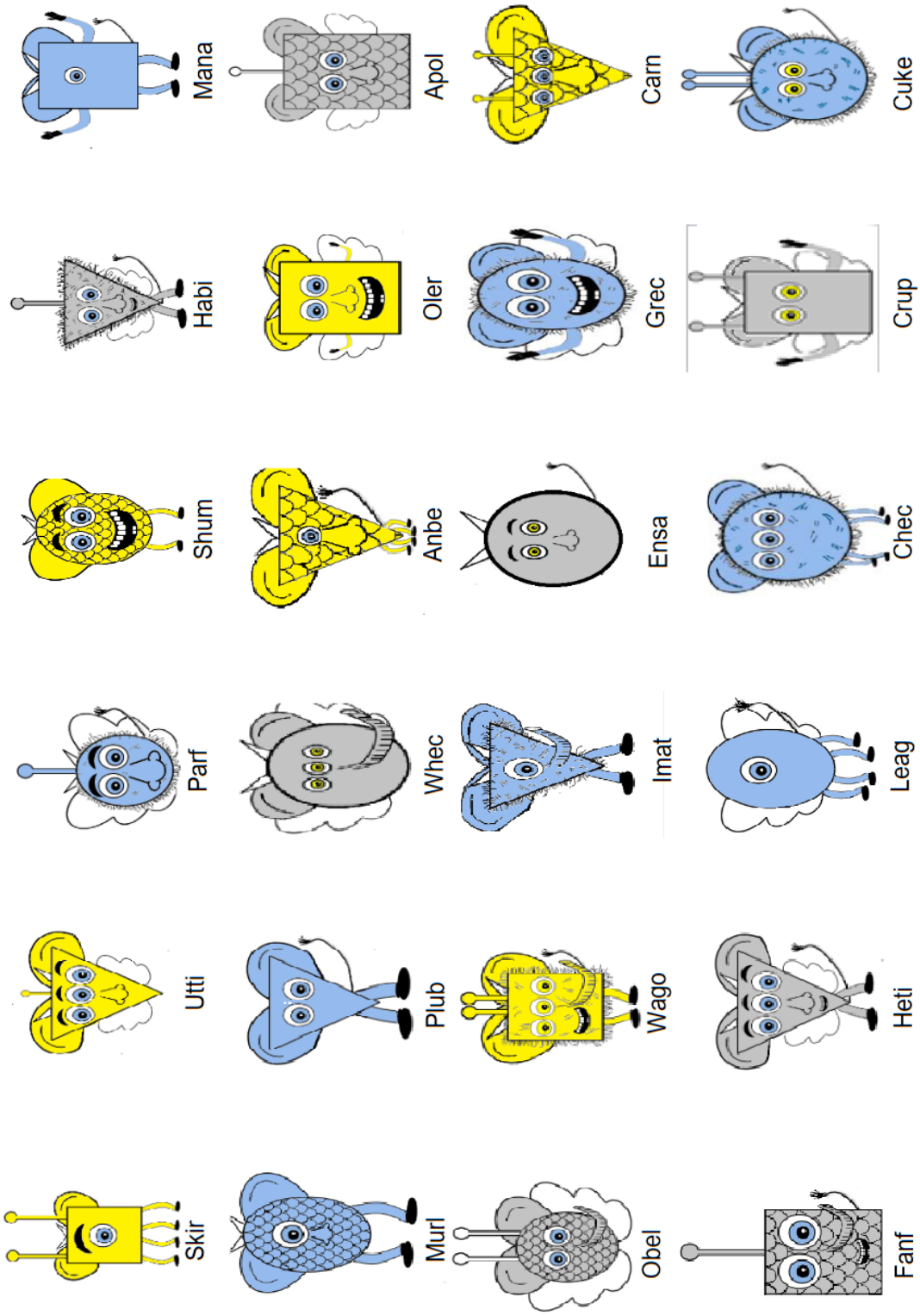
The pupil will engage in short, game-like tasks. Teachers and parents may be asked to fill out some short questionnaires.

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 E: Pictures used in Pavitt (2017)



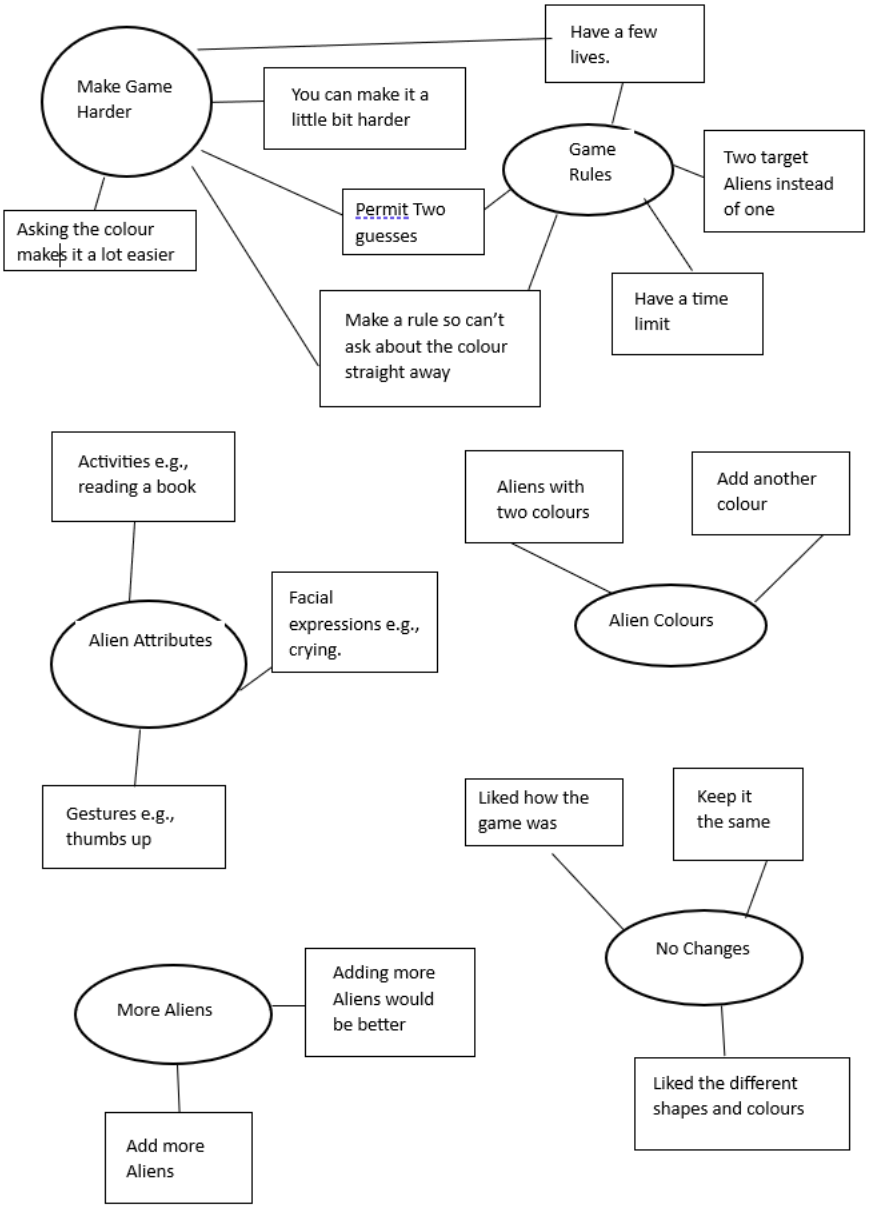
APPENDIX F: List of 13 Alien Attributes Used in Current Study

Alien Attributes	Variations
Horns	Horns
	No Horns
Tail	Tail
	No Tail
Ears	Ears
	No Ears
Arms	Arms
	No Arms
Wings	Wings
	No Wings
Eyebrows	None
	One
	Two
	Three
Teeth	Teeth
	No Teeth
Shape	Circle
	Triangle
	Square
Colour	Grey
	Blue
	Yellow
Antenna	One
	Two
	Three
Eyes	One
	Two
	Three
Nose	No nose
	Small nose
	Big nose
Legs	None
	Two
	Four

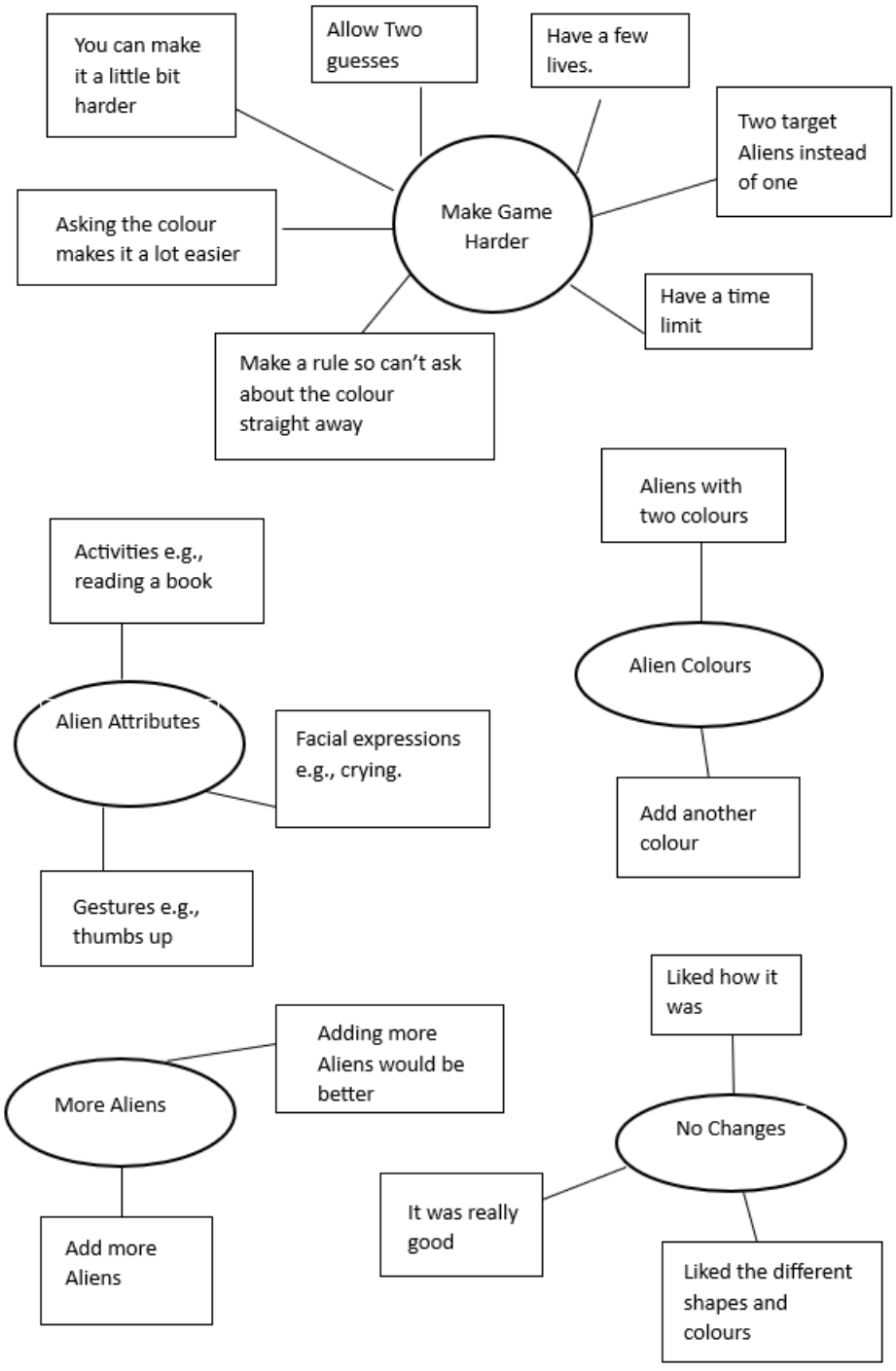
APPENDIX G: Alien with Superordinate Attribute

Alien with a head and body created a separate attribute (i.e., a superordinate attribute); it was therefore removed from final game design.

APPENDIX H: Thematic Mapping of Initial Codes



APPENDIX I: Thematic Mapping of Final Codes



APPENDIX J: University of East London School of Psychology Research Ethics Approval

University of East London

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 Task as an Assessment of Concept Formation 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?
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. Via the school, we will provide information sheets (accessible format for the children) and opt

		in consent forms for the children and carer to read in order to decide whether to take part. Parents are asked to contact us via email if they have any questions about the study. Children will only be able to take part if their parent/carer gives consent and returns the consent form to us via the school. If consent is given by the parents/carer, we will introduce and discuss the study with the child and seek consent or assent as appropriate.
3.7	Measures, materials or equipment: Provide detailed information, e.g., for measures, include scoring instructions, psychometric properties, if freely available, permissions required, etc.	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 (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.		
5.3	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:	YES ☒		
5.4	If necessary, have appropriate support services been identified in material provided to participants?	YES ☐	NO ☐	N/A ☒
5.5	Does the research take place outside the UEL campus?	YES ☒	NO ☐	
	If yes, where?	The data collection will take place on primary school campuses.		
5.6	Does the research take place outside the UK?	YES ☐	NO ☒	
	If yes, where?	Please state the country and other relevant details		
	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	YES ☐		

	survey), regardless of the location of the researcher or the participants.	
5.7	**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). 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		

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 20-10 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 Summer and Autumn school term, from June to November 2022. 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, we 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

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.	2	2	4		4

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:	Using a Game-Like Task as an Assessment of Concept Formation in Children

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 BEFORE THE RESEARCH COMMENCES	In this circumstance, the student must confirm with their supervisor that all minor amendments have been made before the research commences. Students are to do this by filling in the confirmation box 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 confirmation from students where minor amendments were required, must be obtained before any research takes place. For a copy of UEL's Personal Accident & Travel Insurance Policy, please see the Ethics Folder in the Psychology Noticeboard.

Confirmation of minor amendments (Student to complete)	
I have noted and made all the required minor amendments, as stated above, before starting my research and collecting data	
Student name: (Typed name to act as signature)	Pinar Marasli
Student number:	U2075213
Date:	01/11/2022
Please submit a copy of this decision letter to your supervisor with this box completed if minor amendments to your ethics application are required	

APPENDIX K: Organisation 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 (3 days) 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 L: Organisation Information Sheet

ORGANISATION INVITATION LETTER

Using a game-like task as an assessment of executive functions in children.

You are being invited to participate in a research study. Before you agree it is important that you understand what participation would involve. Please take time to read the following information carefully.

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.

What is 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.

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.

Why has your organisation been asked to participate?
Your organisation has been invited to participate in my research as you host the kind of people we are looking for to help us explore our research topic. We are looking to involve children aged 6 to 11.

You, as an organisation, are quite free to decide whether or not to participate and should not feel coerced.

What will your participation involve?
If you agree to participate, the children you host, and their parents, will be asked whether they would like to participate in this study. Children will then attend a session with 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 children for participating in my research, but their participation would be very valuable in helping to develop knowledge and understanding of our research topic. To thank the school for taking part in our

research we will be offering a CPD session for staff, the topic of which can be chosen from a pre-selected list.

Taking part will be safe and confidential

The children's privacy and safety will be respected at all times. Participant's data will be kept anonymous, meaning they will not be able to be identified by the data collected, on any written material or in the write-up of the research. Parent's and children's consent forms will be stored securely and separately from the rest of the data and will be destroyed following completion of the research.

Participants do not have to complete all tasks asked of them and are free to stop their participation at any time.

To ensure the children's and our own safety, social distancing will be maintained at all times, we will wear a mask and sanitizing of hands and equipment will be completed regularly.

What will happen to the information provided?

What we will do with the material children provide will involve anonymously storing all data on a personal drive, only we have access to, which will be password protected. 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. The anonymised data will be reviewed by ourselves and our supervisor and may be requested by examiners. Summaries of the data collected will be available in the write-up and may be published in an academic journal, the thesis will also be publicly accessible on UEL's institutional repository. 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. Once the research has been completed, the data will be kept for three years, following this, the data will be destroyed. Once the data has been collected children and their parents can withdraw the data up to the end of January* 2022.

What if a child or their parent wants to withdraw?

Children and their guardians are free to withdraw from the research study at any time without explanation, disadvantage or consequence. Separately, children and their guardians may also request to withdraw their data even after they have participated data, provided that this request is made before the end of January1 2022 (after which point the data analysis will begin, and withdrawal will not be possible).

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 via email: alien.game@uel.ac.uk

If you have any questions or concerns about how the research has been conducted please 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)

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 M: Parent 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 January2 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)

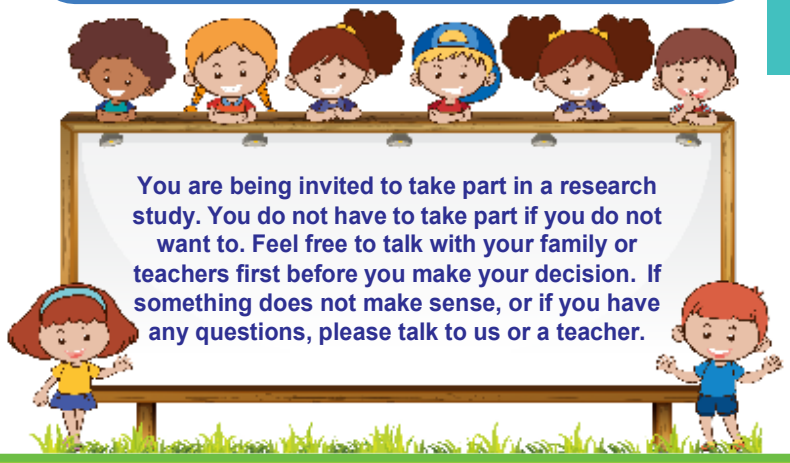
Thank you for taking the time to read this information sheet.

APPENDIX N: Child Information Sheet

Consent to participate in a research study

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

University of East London



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 O: Parent Consent Opt-out 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 survey, (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 survey, 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 P: Child Consent Form

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 Q: Child Debrief Sheet

PARTICIPANT DEBRIEF SHEET

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

University of East London

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 R: Demographics Sheet

Name/ID:
DoB:
Age:
Sex/GI:
Nationality/Ethnicity:
Primary language:
Other language(s):
Sensory or motor needs:
Parent's occupation:
Test by:
Test date:
Test location:
Notes:

APPENDIX S: Response Sheet

Trial A Target Alien:
Game questions asked: (questions that elicit a yes or no response)
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Other questions or queries:

Notes on strategy and/or behavioural observations: (e.g., engagement, distractibility, motivation, task enjoyment etc.):