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

Alexandros Bardis

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

Concept formation is a term that refers to the ability to create abstract categories and rules for the objects in our environment. It is an ability that develops throughout childhood, and is important in education and learning. However, there are limited concept formation measures for six- and seven-year-old children, many of which are adaptations from adult tests. The present study seeks to explore the utility of a game-based task, called the Alien Game, originally designed by Pavitt (2017) as a child-friendly and culturally fairer measure of concept formation.

In this cross-sectional study, 13 six- and seven-year-old participants, completed the Alien Game, and WISC-IV Similarities and Matrix Reasoning. Teachers completed working memory and inhibition ratings. The concurrent validity, applicability and likability of the Alien Game were explored.

The Alien Game Abstraction score (AS) had high applicability as a reflection of children's strategy based on concept formation. Spearman's Rank correlation tests found a good concurrent validity of AS with Similarities and Matrix Reasoning scores, indicating that the AS is a good measure of concept formation. The AS had low predictive validity of working memory and inhibition. The Alien Game was rated highly by participants, who preferred it to established measures of concept formation.

The Alien Game merits further development as a concept formation measure for six- and seven-year-old children. Further development should involve replication with a larger sample, analysis of the relationship between culture and performance in the Alien Game, and clinical utility for children with neurodevelopmental diagnoses and brain injury.

ACKNOWLEDGEMENTS

This is the biggest piece of academic writing that I have ever undertaken. It is the culmination of my clinical psychology training and a journey of many years. I am not going to be able to do justice to all the people that have contributed to my work in big or small ways.

However, it is important that I thank my supervisor, Dr Matthew Jones-Chesters, for your support, guidance and advice. Thank you in particular for being there during the tricky times of data collection and analysis!

Thank you to my Alien Game research colleagues, Emily Hay, and Pinar Marasli. I enjoyed working with you in the making of the Aliens and thank you for your support and input. Thanks is due to Annie Pavitt, who piloted the first Alien Game.

Thank you to the school and all the children who agreed to take part in this study and engaged with the Alien Game. The school staff team were very supportive and welcoming, and thank you to the teacher for filling out all those questionnaires.

A very big thank you to my partner, Jack, and my friends Maria and Natty for their emotional and practical support and cheerleading while I was busy writing. A big thank you to my mum and dad, Poppi and Philippos for their infinite support and love; to Olwen and David for their support and hospitality; and all my family, friends, colleagues and supervisors who have been there for me throughout the years. Thank you to all, from the bottom of my heart!

Finally, I want to thank Lydia, my life coach, for helping me stay organised in the last two years of training, and Toulis, my last remaining brain cell, for not giving up!

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

“If the human brain were so simple

That we could understand it,

We would be so simple

That we couldn’t.”

G.E. Pugh (1977, p.154)

1.1 Overview

The study and measurement of human cognition is one of the main areas of clinical psychology. It is an endeavour ripe with theories and debates, split into numerous theoretical, epistemological, methodological, and experimental directions. The lack of a unifying theory of cognition is a mere reflection of how complex, intricate, and multifaceted the mind is.

The present study belongs to this body of research. It aims to address the development of a novel assessment tool for children, designed to test concept formation, which is a subset of the executive functions. First piloted by Pavitt (2017), this novel test is called the Alien Game. The Alien Game aims to be a measure of concept formation that has been specifically designed for children, and thus has a game-based format. The second main feature of the Alien Game is that is seeks to reduce testing biases that can disadvantage children from minoritized cultural backgrounds.

In the following sections of the introduction, more detailed information will be presented on the key concepts of the present study. Firstly, a review of the main theoretical concepts will be offered, namely of executive function, concept formation, and cultural fairness. Secondly, the findings of a literature review on existing concept formation assessment measures (for six- and seven-year-old children) will be discussed. Finally, the work of Pavitt (2017) will be reviewed. It will be explained how the present study seeks to extend the Alien Game and the aims and research questions will be outlined.

1.2 Theoretical Introduction of Executive Functions and Concept Formation

1.2.1 Executive Functions

Executive function is an umbrella term for higher order cognitive abilities. No consensus has emerged yet as to how best to define executive functions, nor their components. Some researchers posit that the main facets of executive functioning are working memory, the ability to understand and use rules, planning, abstract thinking, flexible thinking and error evaluation (Jacques & Zelazo, 2001). Other researchers argue that additional abilities belong to the executive functions group, including ability to sustain focus and attention and the mental manipulation of information (e.g. Rennie et al., 2004). The terms to describe each ability can vary. For example Jacques and Zelazo's flexible thinking can be referred to as fluency or inhibition (Lee et al., 2004). In contrast, for Miyake et al. (2000) inhibition and flexibility are distinct components of their Three-Factor Model of executive function. For Miyaki et

al. (2000), inhibition refers specifically to the ability to supress a prepotent response, which is very similar to self-control (e.g. Mischel, 2014), whereas flexibility refers to the ability to apply a new set of abstract rules in response to changes in a situation. The ability to mentally manipulate information described in Rennie et al.(2004) is very similar to working memory, and specifically the episodic buffer and central executive components of working memory (Baddeley, 2002).

The frontal lobe is thought to be the 'seat' of executive functions in the brain (Welsh & Pennington, 1988). A more detailed understanding of the localisation of executive functions in the frontal lobes has led to a distinction between hot and cool executive functions (Zelazo & Muller, 2002). The main distinction between these types is whether a situation bears particular motivational or emotional significance: if it does, then the ventral and medial pre-frontal cortex have been shown to activate; when it does not, the dorsal pre-frontal cortex is activated (Zelazo & Muller, 2002).

Executive functions were traditionally believed to be unavailable to children prior to the maturation of the frontal lobes in early adulthood (Golden, 1981). However, once developmentally appropriate adaptations of executive function tasks were developed, it was soon found that children too utilised executive functions, but that these matured in different stages (Passler et al., 1985). This prompted an understanding that some executive functions are available to children from very early in their development, albeit in a very preliminary form (Diamond, 1988).

Overall, executive functions have a strong future orientation, as one of their purposes seems to be to help an individual formulate a goal, and make a plan of how to achieve it, with cognitive constructs such as working memory

and inhibition standing out as the most important components (Lee et al., 2004). One implication is that although the various executive functions may be demarcated, in practice they support each other in tasks that require higher-order thinking. However, the lack of a clear definition and clear developmental framework, has led to inconsistencies in the way executive functions are measured and researched (Lee et al., 2004).

1.2.2 Concept Formation

Bruner et al. (1967) define concept formation as the ability to categorise entities based on their shared characteristics. Our ability to categorise the world is central to our ability to make sense of it (Bornstein & Arterberry, 2010). Concept formation is an ability particularly important for education, as it has been found to correlate with children's school performance (Gligorović & Buha, 2013). More specifically, it is important in learning to be able to acquire new concepts and use them in a flexible way to formulate different conceptual strategies to a learning problem (Van der Sluis et al., 2007; Bull & Scerif, 2001). It is thus important to have reliable tests that help in the early detection of developmental delays, so support can be put in place for children who need it. Hence the need for research on reliable concept formation assessment tools for younger children.

As with executive functions, there is little consistency in the definition of concept formation. It is closely related to abstract thinking, which is one of the constituent abilities of executive function presented above (Goldstein et al., 2014). Abstract thinking is a broad term that refers to what Siegler (1991) calls Conceptual Reasoning. This is the ability to carry out cognitive

processes around the perception and representation of abstract concepts. Furthermore, it is the ability to use these concepts in a flexible manner. This requires the ability to inhibit concepts that have been previously used but are no longer relevant. Indeed, inhibition appears to play an important role in concept formation, as will be discussed later. Concept formation is, therefore, only a part of conceptual reasoning. It can be understood as the ability to perceive salient features, and to extract increasingly more abstract concepts out of the sensory input and information available in one's environment. Another term to describe concept formation, and more specifically its earlier developmental form more relevant for infants and toddlers, is Categorical Learning (Condy et al., 2021).

1.2.3 Development of Concept Formation

Concept formation can be based on perceptual processes, and this is defined as perceptual concept formation (Condy et al., 2021). Perceptual concept formation involves the processing of sensory input from physical properties, e.g. being able to visually determine if two entities have the same colour (physical information may include the way an object functions or the way it is used). An example of this type of concept formation is illustrated in a visual habituation study by Booth et al. (2010): they demonstrated the ability of 14–18-month-old toddlers to form novel conceptual categories. The toddlers were shown novel objects which were used in a specific way to demonstrate their function; the toddlers proved able to form conceptual groupings of these novel objects with respect to their function.

Function facilitates the formation of thematic relationships across objects (i.e. the way they are used, Perraudin & Mounound, 2009), and this plays a part in concept formation until approximately the age of seven years. These authors showed that unless a novel object can be associated with an action (what the object does or how it is used), it will be difficult for a child to form a conceptual representation of that object. There seems to be a developmental shift at the age of seven years from thematic relationships (e.g. bread and knife go together because the knife is used to cut the bread) to more abstract categorial relationships (bread and cake go together because they are both baked food items, Perraudin & Mounoud, 2009). This second type of concept derivation is called semantic concept formation (Condy et al., 2021). Semantic concept formation refers to the capacity to use linguistic representations (e.g. determining whether an object belongs to the category "blue entities".).

Several theories attempt to explain the developmental shift from thematic to semantic concept formation. These theories focus on the roles of procedural and conceptual knowledge development that may support this shift. These refer to knowing "how" things can be used, versus understanding "why" things work in a certain way, respectively (Schneider et al., 2011).

Karmillo-Smith (1992) suggests a Procedures-first model, meaning that the acquisition of procedural knowledge is the foundation for the emergence of conceptual knowledge. The obverse is suggested by Gelman and Williams (1998): in their Concepts-first model, they posit that children must have the conceptual basis available to understand a task, and get better at it with practice. Both of these two models are based on the premise that one type of knowledge exists first and supports the emergence of the second.

In the Inactivation model (Hasspasalo & Kadjievich, 2000), the development of procedural and conceptual forms of knowledge are taken to be independent of one another. Finally, there are iterative models (e.g. Rittle-Johnson et al., 2001), which suggest that other than there being a uni-directional influence between procedural and conceptual knowledge, they instead influence each other in a bi-directional fashion. In their study, Schneider et al. (2011) assessed the direction of the influences of procedural and conceptual knowledge and found more evidence in support of an iterative model. In addition, they assessed the effects of prior task knowledge and found that this had no moderating effects on the direction of influence. Schneider et al. (2011) argue that conceptual knowledge is necessary for a child to be able to evaluate the method they use to problem-solve in a given situation, which is referred to as procedural flexibility. They suggest that procedural flexibility helps to focus attention on the most relevant aspects of a task.

Concept formation development may be different for children with diagnoses of autism (Naigles et al., 2013), ADHD (Lee et al., 2004), learning disabilities (Campbell et al., 2013) and brain injury (Canfield et al., 2004), therefore the theoretical understanding of concept formation presented above may not apply in the same way. However, this is beyond the scope of the present study and therefore will not be reviewed here.

1.2.4 Role of Working Memory in Concept Formation

As mentioned above, working memory and inhibition are both important executive functions, and they both appear to play an important role in concept formation. According to Hayes et al. (2013) the ability to detect similarities between objects plays a significant role in the development of concept formation in children, especially before the age of seven years. However, this requires the ability to remember objects previously seen, in order to be able to make comparisons, and detect degrees of similarity and difference, that can then lead to development of further concepts, categories and taxonomies (Hayes et al., 2013). Dauvier et al. (2014) posit that children between the ages of five and eleven years are able to process up to three conceptual characteristics at any one time.

Even though working memory seems to be important in supporting the development of concept formation, it is not clear how the two interact to influence performance during a task. Some theorise that substantial working memory demands will interfere with concept formation (Crone & Van der Molen, 2004); therefore a task with high working memory demands will not reliably measure a child's concept formation ability (McGee et al, 2008). In contrast, McGonicle-Chalmers and Alderson-Day (2010) suggest that conceptual rules and taxonomies emerge when the load on working memory is high.

It is possible that these accounts reflect different types of working memory. For example, McGonigle-Chalmers and Alderson-Day (2010) refer to Prospective working memory as a mechanism that enables the child to store alternative ways of problem-solving action until they can compare them and decide which one might be most successful. In that sense, this increases

pressure for a grouping heuristic (Principled sorting heuristic) where the child creates groups of objects in order to store them better in their working memory. This is very similar to what Phillips et al. (2014) refer to as an automatic pathway of categorical learning.

1.2.5 Role of Inhibition in Concept Formation

Inhibition plays a significant role in concept formation. Conceptual inhibition is a specific type of inhibition that enables categorical flexibility (Blaye & Jacques, 2009). Flexible use of concepts is enabled by preventing the perseveration of irrelevant ones. Categorical flexibility is similar to the ability to switch between categorical rules and abstract rules. Kharitonova et al. (2009) make the distinction between switchers and perseverators to refer to groups of children who have broadly developed the ability to switch between rules and categories, and those children who have not. Switching is achieved typically by the age of four to five years (Kharitonova et al., 2009).

A proposed model to explain this difference between switchers and perseverators is the Cognitive Complexity and Control (CCC) Theory (Zelazo et al., 1996). According to the CCC, perseverance is the due to the lack of a superordinate rule. A superordinate higher level rule (e.g. “I can use either rule A or rule B”) enables a child to switch between the rules that that fall under the superordinate rule.

Van Bers et al. (2014) discuss additional reasons why perseverance may happen. One possibility is that perseverators understand that there is a new rule, but find it difficult to shift their attention to it. This is referred to as the Attentional Inertia Theory (Kirkham et al., 2003). The Competing Memory

System models (Cohen & Servan-Schreiber, 1992; Morton & Munakata, 2002) is a slightly different version of this scenario, and involves memory. In this model, the memory-representation of the new rule is not as strong as the old rule and therefore fails to be applied.

Another explanation is the Activation Deficit Model (Chevalier & Blaye, 2008; Müller et al., 2006). According to this model, perseverance occurs because the new rule was inhibited when the old rule was activated. A previously inhibited rule will thus be difficult to activate at a later time. A fourth possibility is that the use of a new rule requires the reconceptualization of the task, and this is known as the Redescription Account (Perner & Lang, 2002).

1.2.6 Theoretical integration of Concept Formation

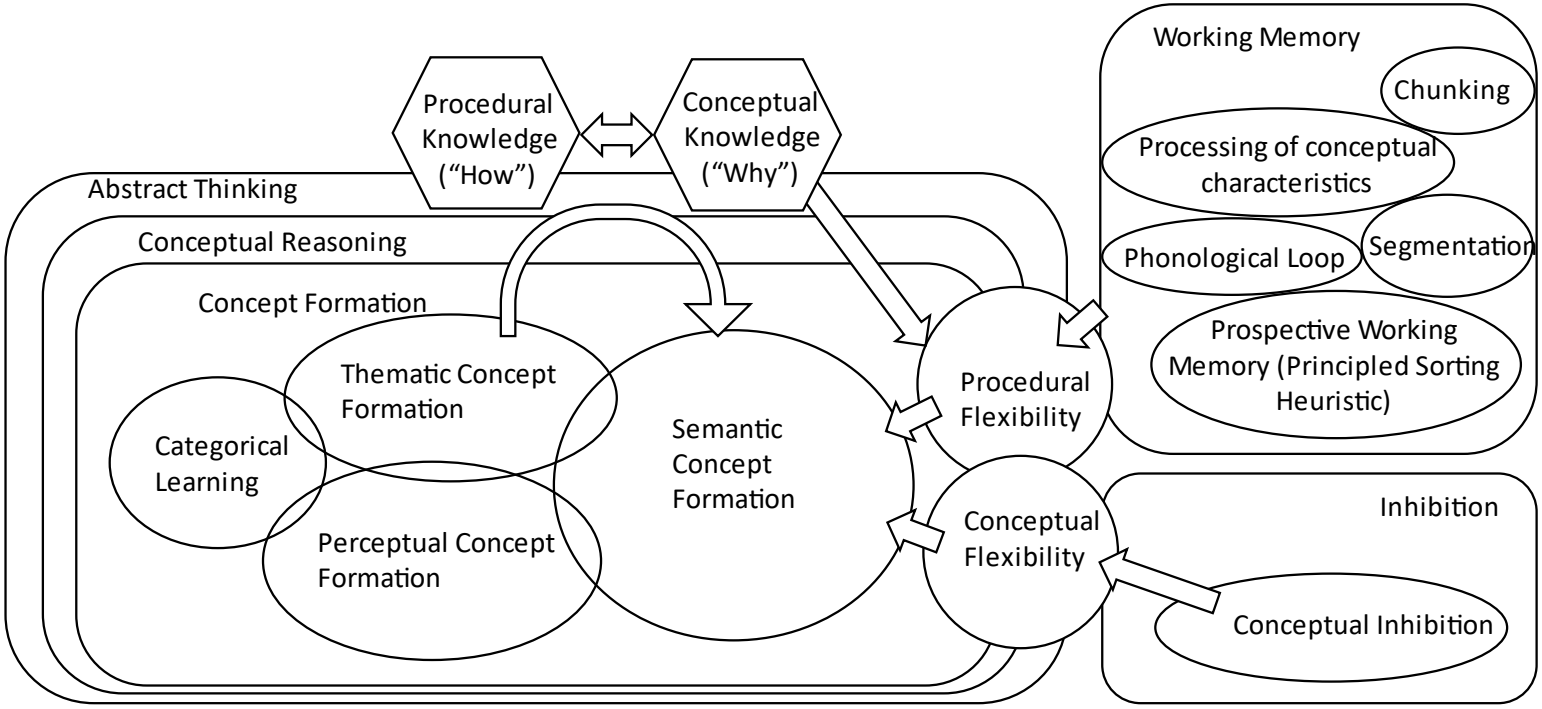
There is no consensus cognitive model of concept formation. The sections above have demonstrated how different authors use different terms to describe abilities that fall under the wider umbrella of abstract thinking and, within that, conceptual reasoning.

As illustrated in Figure 1, concept formation appears to encompass a number of diverse sub-abilities. Categorical learning is associated with early development and overlaps with perceptual concept formation, since both refer to the categorisation of objects in terms of physical properties. The accumulation of conceptual and procedural knowledge leads to greater abstraction in concept formation, and facilitates a shift from thematic to semantic concept formation around the age of seven years. Working memory supports concept formation through procedural flexibility, meaning that a number of different ways a situation can be conceptually formulated and

problem-solved can be held in mind and evaluated. Inhibition supports this process through conceptual flexibility, meaning that each alternative conceptual formulation in that situation can be inhibited so that another one can be applied, especially if the previous one was not effective or when the situation changes.

Figure 1.

Theoretical integration of Concept Formation, Working Memory and Inhibition



1.2.7 Factors that influence Concept Formation performance

Additional factors contributing to the lack of clarity on this matter are derived from methodological limitations in concept formation research. These are related to the structure of the tasks, and whether they intersect with working memory and inhibition demands.

Inhibition may not only be helpful in preventing an old rule or concept from being applied. Motor inhibition is required in concept formation tasks where a child needs to perform physical actions to indicate their responses. For example, in a card sorting test, Kirkham et al. (2003) demonstrated that children could obtain better scores if they named the cards that they wished to sort as opposed to physically having to pick them. The physical sorting action was found to be prone to motion perseverance, reflecting the action the children had performed previously. Greve et al. (2000) suggest that another way of reducing the motor needs in a task is through the handling of the materials and cards by the person administering the assessment rather than the child, after the child has indicated their response.

Generally, any reduction in the need for motor responses in a test will help reduce contextual biases that may influence a child's performance (Blay & Jacques, 2009; Condy et al., 2021). For example, sometimes instructions or labels on stimuli may not be clear or salient enough for child participants.

Test length may also affect a child's performance. Canfield et al. (2004) reported that, where participants found tests too long, this caused frustration. Level of difficulty could be another factor that impacts on frustration (Canfield et al. 2004), as can anxiety. An artificial testing environment or anxiety

provoking task could interfere with a child's ability to perform (Keller, 2021). It is therefore highly important to help a child feel at ease with the task format, so a game-like structure may be an effective solution to some of these problems.

Condy et al. (2021) recommend a move away from more traditional pen and paper cognitive tests, and encourage the development of computerised tests, and the use of touch screens, as this may make them more accessible. The mode of testing and recording responses, however, may introduce another element of complexity. Condy et al. (2021) reported that participants found it hard to manipulate items on a touch screen in order to categorise them, thus highlighting the importance of considering user-friendliness in computerised tests.

A computerised test may not be as valid or accessible for children who have grown up in a digitally restricted environment (Kaarakainen & Saikkonen, 2022). As digital exclusion affects people from lower socioeconomic strata (Park, 2017), it is important that cognitive tests do not introduce more bias against already disadvantaged children and thus perpetuate discrimination. It may be more appropriate for a cognitive measure to not require either pen and paper (i.e. writing) or use of advanced or inaccessible technology.

1.2.8 Summary

It is important for concept formation assessment tools to have a design that does not tax working memory nor demand conceptual inhibition. In addition, responses should be given in a way that does not introduce errors due to motor perseverance. Verbal input should be kept to a minimum, and any

verbal instructions given are as clear as possible. Length of the tasks should be carefully considered to avoid inducing frustration. To minimise anxiety, the child taking the test would benefit from an administration method that is not too interrogative; they would also benefit from being in a familiar environment. Finally, it seems best to avoid tasks that demand literacy or technical skills, as different groups of children may find this inaccessible.

1.3 Cultural fairness in cognitive assessments and research

An important consideration for any psychometric instrument is whether it introduces or mitigates biases that may affect performance and therefore disadvantage certain groups of people being assessed. In the previous section it was noted how task administration and design may introduce psychological and cognitive-related confounding variables in a person's performance. In this section, it is shown how social and cultural biases may affect performance.

The importance of cultural fairness has to be understood in the context of the history of cognitive assessments. The emergence of intelligence and cognitive testing in the Western World was later followed by questionable interpretation of cognitive scores, for example of Black people, compared to White people, as objective evidence of their biological inferiority (e.g. Garret, 1947; Gottesman, 1968). This is just one example of the damaging consequences of epistemic oppression (Buchanan et al., 2021) and support for racist policies.

Meta-analyses have shown that the differences in cognitive assessment scores in adults are approximately one standard deviation in magnitude between White people and Black people, and two thirds of a standard deviation between White people and Latin American people (Roth et al., 2001). These results are based on racial taxonomy of the population. Racial taxonomy is very poorly defined in the literature, and usually serves as a proxy for culture and socioeconomic status (Keita et al., 2004). Therefore it is

important that these differences are approached from a socio-cultural perspective.

Culture can be defined as “learned or acquired behaviours or traits attributable to the socialization experiences resulting from membership in particular systems or institutions within a society” (Helms, 1992, p.1091). In an increasingly globalised world there is need to understand that culture may not be static but increasingly diverse and fluid (Fernandez & Abe, 2018). This may be particularly relevant to areas of high cultural diversity such as parts of London in the UK. In areas such as these there is a need for cognitive assessment tools to be available that do not disadvantage marginalised groups of the population due to the cultural bias of the tool itself.

The influence of culture may be profound. Firstly, intelligence itself may be conceived in different ways from one culture to the next. For example, the Kenyan concept of intelligence deviates from the Western conceptualisation, as it expands beyond the territory of knowledge and ability to include aspects such as respect (Grigorenco et al. 2001). Secondly, there is evidence from neuroimaging studies that culture can lead to a level of differentiation in the way different cognitive tasks engage different parts of the brain (Fernandez & Abe, 2018). When considering concept formation, culture is a very important determinant of the types of concepts that will become available to a person growing up, through which they will understand their environment (Carey, 2009).

Research on the influence of culture has led to a better understanding of the main factors that introduce cultural bias in cognitive assessments: construct, method, and item biases.

Construct bias may be introduced when a test lacks construct validity, i.e. that there is a mismatch between what a test is intended to measure and what actually is measured. According to Bryman (2008), there are two ways construct validity can be affected, both related to the quality of the definition of the construct a test aims to measure. The definition can either be too narrow, so that it misses aspects of the construct, or it can be too wide, and captures extraneous elements that are not relevant (Helms-Lorenz & Van De Vijver, 1995). As an example, verbal fluency (word generation) was once considered an index of verbal output, but is now recognised to principally involve executive functions (e.g. task switching).

Item bias is the way specific items within a test influence a person's performance, and has been the focus of much research. Such bias can be introduced by cultural familiarity, i.e. that an item may be more or less accessible depending to a person's exposure to the culture or environment associated with that item (Helms, 2002). This may also happen due to different social factors within a single culture, such as socio-economic status and class (Helms, 2002). Finally, language plays a very important role in item bias, especially when tests are translated to be used cross-culturally. Items translated into a new language may not be understood in the same way in the new cultural context (Helms-Lorenz & Van de Vijver, 1995).

The standardisation, or the test administration, may be prone to cultural loadings (Van de Vijver & Poortinga, 1992), which can result from a number of different factors. The cultural identity of the test administrator, the cultural identity of the person being tested, and the interaction between the two are all significant factors. Cultural differences, as well as difference in the primary language spoken, may lead to cultural discrepancy. Test-wiseness, i.e. a

person's familiarity with the overall testing procedure and understanding of testing conditions and appropriate behaviour in that context, is an additional factor (Sarnacki, 1979). This is related to pre-existing familiarity that the person being assessed may have to a particular item, stimulus or test. In essence, tests have the tendency to measure a group's exposure to the dominant culture, as it is the culture that has influenced the nature and structure of the test. Groups from the non-dominant culture may be affected differently by cultural loadings, dependent on their own acculturation strategy (Berry, 1994).

According to Anastasi (1968), the defining feature of a culturally fair test is that it has the ability to control for the cultural biases outlined above. However, Arvey (1972) argues that this is an ambitious definition. To meet this criterion we need to be certain that we understand all factors that underpin performance, that we understand how they influence performance in that specific test, and that we have devised the statistical means to mitigate, or even eliminate these factors. Even then, this process needs to work for all cultures.

The inverse approach is used by Darlington (1971), who argues that it is possible to have a culturally fair test when it is proven to not be influenced by any culture. In other words, the test scores do not correlate with the indices. This is very similar to Meredith (1993), who argues that if factor loadings or regression weights of a test are shown to be equivalent across cultural groups, then the scores of people from different cultures could be meaningfully compared. This is one of the ways that procedural equivalence can be achieved, i.e. that the test's methodology does not introduce cultural bias (Johnson, 2006).

Alternatively, interpretative equivalence (Johnson, 2006) is achieved when the test content and administration is accessible to both the person administering an assessment, and the person being assessed, in a way that does not introduce cultural bias (Johnson, 2006).

1.4 Concept Formation measures for Young Children

1.4.1 Literature Review Method

There are few widely available concept formation tasks for young children, e.g. aged six and seven years (Smidts et al., 2004). A downward extension of the Alien Game seeks to:

- add a game-based concept formation measure to existing tests
- utilise a design that increases its child-friendliness
- use culturally fairer stimuli to mitigate cultural factors that may influence a child's performance.

The following literature review assesses the existence of similar game-based measures.

Combinations of keywords were used to search databases of peer reviewed articles. The keywords used were: (Concept formation or Abstraction or Induction) and Child* and Test* and Neuropsych*). The databases searched were PsychInfo, CINAHL complete, EBSCO Host, Child Development and Adolescent Studies, Psych Articles.

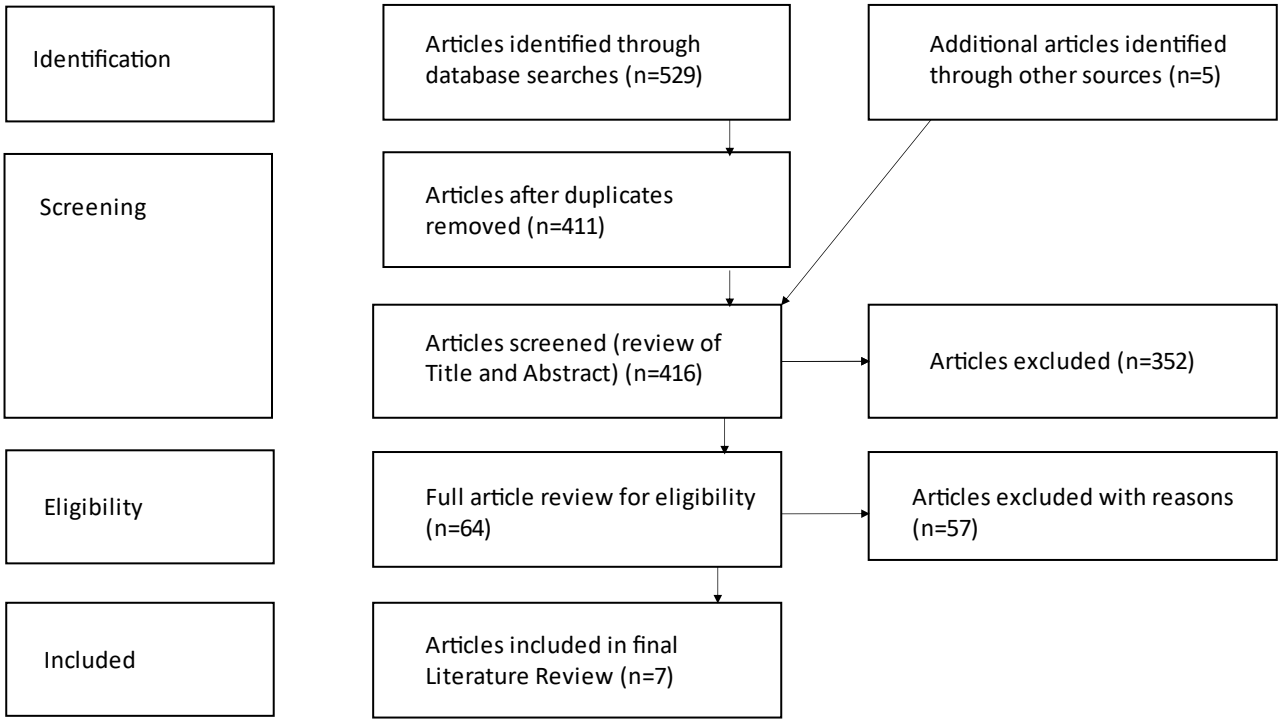
This yielded a total of 533 articles, which were then screened for relevance according to the following exclusion criteria (see Fig. 2).

Items were included in the final literature review sample if they described a method of assessing concept formation in children aged six and/or seven years. Items were excluded if the sample of children had a specific neurodevelopmental diagnosis or brain injury, if the study was not in a peer

reviewed article, if it was not written in English, or it was not accessible for full review.

Figure 2.

PRISMA (2009) Flow chart for Article Selection Process



1.4.2 Existing Tests and Instruments

1.4.2.1 *Raven's Standard Progressive Matrices (SPM)*

The Raven's Standard Progressive Matrices (SPM; Raven et al., 2000) is a non-verbal task that works by presenting the participant with a series of visual patterns. Each visual pattern has a gap, and the person completing

the test has to choose which option out of multiple choices will correctly complete the pattern. The items become progressively more difficult and require higher levels of working memory to manipulate information and identify the rule or pattern.

The SPM can be used with children aged five years upwards, and is a standardised measure. This means that a person's score on the SPM can be interpreted in terms of its closeness to the average performance of their age group. The SPM has been standardised in many countries (Raven, 1989), demonstrating cultural awareness of the test. In fact, the visual stimuli used for SPM are thought to be relatively culture-free, therefore more accessible and less likely to introduce cultural bias (Raven, 1989).

Traditionally, it has been assumed that the SPM is underpinned by a general factor of fluid intelligence (Kunda et al., 2016). However, research indicates that people are applying different methods of problem-solving to the SPM, raising questions about the test's construct validity (Kunda, et al., 2016). Overall, the SPM is one of the few standardised measures that can be used for children aged six and seven years old.

1.4.2.2 *Feed the Hungry Donkey*

Crone and Van der Molen (2004) designed a task to measure inductive thinking in children (inductive is a form of abstract thinking related to concept formation). Feed the Hungry Donkey is a computerised game-based adaptation of the Iowa Gambling Test for adults (Bechara et al. 1994), where participants had to decipher which strategy would result in highest gains. Opening a door in the game would lead to gaining as well as losing apples

for a fictional donkey. Participants were required to select one of four doors to open in order to retrieve the most apples for the donkey. Two doors involved an initial high reward followed by significant losses and two other doors led to small gains followed by smaller losses. The task was designed to measure concept formation skills relating to the ability to deduce the rules of the game, and to group the strategies as high versus low risk.

Participants were divided into age groups, six to nine years, ten to twelve years, thirteen to fifteen years and an adult group of eighteen to twenty-five years. Performance yielded gradual improvement in the number of low risk choices for all age groups. Six- to nine-year-olds showed the least improvement and eighteen to twenty-five year-olds made the highest improvement. In fact, Bechara et al. (1997) argue that the Iowa Gambling Task is a measure of conceptual knowledge, and it is not known whether it lends itself for a downward extension for children. Downward extensions run the risk that, due to developmental processes, they measure a much less mature, or even different, set of skills. Therefore they may lack the sensitivity and construct validity needed to be used with children (Bello et al., 2008). However, the game-based design of the test may have increased its child-friendliness.

1.4.2.3 *Relational Integration Level Assessment Task (RILAT*)

Dauvier et al. (2014) designed an assessment tool addressing *relational integration*, which they defined as the ability to detect similarities between objects and then abstract these relations into concepts. They drew upon and adapted previous tests, such as the Relational Matching Task by Bunge et

al. (2009) and the Latin Square Task by Birney et al. (2006), which place high demand on working memory capacity. Their aim was to create a test especially appropriate for children younger than 11 years, since it is difficult to process multiple relational variables simultaneously prior to that age (Andrews & Halford, 2002).

In RILAT, there are visual objects that can have up to four characteristics: different colours, shapes, texture, and numbers of stars inside them. The person assessed is presented with twelve items, where one, two or three objects are shown at a time, on a screen. The task is to select an object, from multiple choices, in order to match various stimuli. The selection rule is that the matching object may only share one characteristic with each of the presented objects. For example, if a blue circle is presented, then the matching object must be either blue in colour but not circular, or circular but not blue. If a striped blue circle and striped red square are presented, then the matching object must match only one of the characteristics of the blue circle and only one different characteristic of the red square. In this example, a blue square or a red circle would match, but a blue triangle for example would not, because none of the presented shapes were a triangle. As the number of presented objects increases, more constraints are placed on which properties the matching object must have to successfully be matched.

Dauvier et al. (2014) used the RILAT with participants aged five to twelve years. They reported that six and seven year old participants performed differently in the task. In the binary items both groups had a high accuracy: 91% for six year olds and 97% for seven year olds. However, in the tertiary items these scores dropped to 22% and 59% respectively. This dropped further for quaternary items to 11% and 25% respectively. These patterns

were analysed to determine if score improvements followed a developmental trajectory. Their interpretation was that children establish relational integration abilities for three objects around the age 7.5 years and for four objects around the age of 10.5 years.

Dauvier et al. (2014) measured the relationship between RILAT and SPM scores, and found a strong relationship between the two. They argued that performance on the RILAT improves when fluid intelligence is higher, because higher fluid intelligence is taken to underpin or facilitate working memory. However, as discussed above, it is not clear what exactly the SPM measures, therefore it is not known what role it may have played in facilitating working memory. Secondly, the authors did not include a working memory measure in their design, which would have been necessary in establishing a link between RILAT and working memory. Moreover, the rules of the RILAT are very complicated, and it is not clear how child-friendly they are.

1.4.2.4 *Wisconsin Card Sorting Test (WCST)*

The WCST (Heaton, 1981) consists of the person being assessed sorting cards that have a number of physical features (e.g various shapes, in various numbers) printed on them. As the participant is sorting the cards, they are given feedback, and based on that feedback they have to decipher the sorting rule. The rule changes periodically and then the person must adapt to the change and decipher the new rule. The WCST can be used with children aged six years five months to adults aged 89 years. Izik Taner et al. (2011)

used the WCST to compare children with a diagnosis of OCD to a control group with no diagnosis to see if there were any differences in abstraction.

The WCST has been critiqued for yielding only a global score of the number of categories achieved, as opposed to the additional scores that are provided with other tests. Moreover, it is not clear what this global score actually measures (McGonigle-Chalmers & Alderson-Day, 2010). The manual responses needed for the WCST make it liable to confounding variables such as motor perseverance and the demands on conceptual and procedural flexibility, meaning that inhibition and working memory are involved (Phillips et al., 2014).

1.4.2.5 *California Card Sort Test (CCST)*

Greve et al. (2000) used the California Card Sort Test (Delis et al., 1992) with participants aged seven to twenty-two years. They were interested in the developmental trajectories in concept formation abilities by children. The CCST can produce a score that reflects three main abilities that may reflect concept formation: concept recognition, concept articulation, and sorting behaviour. These are measured through the three conditions of the test: Free Sorting, in which the child is asked to sort the cards based on their own sorting rules (which they need to state explicitly); Structured Sorting, in which the child observes the examiner sorting the cards, and has to discern what the sorting rule is; and the Cued Sorting condition, in which children are given sorting rules by the examiner and must use those rules to sort the cards.

The CCST cards are divided into three sets of six, with a single word printed on each, which is used for some of the sorting rules (that require verbal ability). Other sorting rules involve the physical properties of the cards, such as shape or colour. The main scores produced by the CCST indicate how many categories a child is able to derive for themselves, how many times they identify a correct sorting rule, and how many times they are able to apply a sorting rule correctly.

Participants were groups of 7 - 9 year olds, 10 - 12 year olds, and 17 - 22 year olds. The results showed that CCST scores improve with age. Notably for the 7 – 9 year olds, their scores in structured scoring were lower than their ability to abstract or apply sorting rules. The authors explained this by theorising that 7 - 9 year-olds find it difficult to explain the sorting rule, which illustrates the contribution and dependence upon verbal ability.

Greve et al. (2000) did not include any children of less than seven years to their sample, therefore it is not known how well younger children would have performed in this task. As the CCST requires participants to manipulate the stimuli, motor perseverance may influence performance. As discussed above, inhibition and working memory are particularly important in card sorting tasks, therefore it is not known how well the CCST reflects only concept formation. The authors did not include any established measures of concept formation in the design for evidence of concurrent validity.

1.4.2.6 *Object Classification Task for Children (OCTC)*

Smidts et al. (2004) developed a measure of concept formation for children aged between three and seven years. They combined and adapted two

existing tests, The Concept Generation test for adults (Levine et al, 1995), and the Concept Generation test for children (Jacobs et al. 2001).

The OCTC involves children sorting toys into groups according to three main variables: colour (red or yellow); size (big or small); and functional use (car or plane). The number of toys could be four or six, depending on the child's ability to understand the instructions and respond correctly to the practice trials. Similarly to Greve et al. (2000), three test conditions were used: free sorting, where the child can sort the toys any way they like; structured conditions where they need decipher the rule the experimenter used to group the toys and state what it was; and the cued condition in which participants group the toys according to explicit rules provided by the examiner.

The results showed that that scores exhibited a linear increase as a function of age, and two developmental phases were suggested: one between three and five years, and another one between five and seven years, (the former associated with the highest increase in performance). They attributed the improvement in performance after five years to increased cognitive flexibility. However, even the oldest children in their sample seemed to find it difficult to shift to a third grouping variable.

This task faces similar challenges to the CCST discussed above. The authors did not include any established measures of concept formation to check its relationship with the OCTC scores. Due to the manual component, the OCTC is prone to the confounding motor function issues discussed above. As a result, it is not clear whether the OCTC is a pure enough measure of concept formation.

1.4.2.7 *Wechsler Intelligence Scale for Children (WISC)*

The WISC (Wechsler, 1974) is a standardised cognitive measure, comprising a number of cognitive subtests. The Similarities and Matrix Reasoning subtests are the ones most relevant to concept formation. In the Similarities subtest, pairs of words are given to the examinee, who has to state the way in which the two words are related, e.g. that red and blue are both colours. The important thing is to state the similarity that is of a superordinate or a higher order. For example, while a cat and mouse both have tails, it is more pertinent to connect them as animals or (even better) as mammals. Similarities is considered a measure of verbal concept formation (Chiappedi et al., 2018), as it requires a semantic understanding of the stimuli.

The Matrix Reasoning subtest is very similar to the SPM. The examinee is presented with visual arrays, one at a time, in each of which there is a gap; given below is a selection of possible responses, and the correct one must be selected. The person being tested can indicate their answer verbally, or by pointing at it. As it is a predominantly visually-based task, it is considered to be a measure of non-verbal concept formation (Brookman-Byrne et al., 2019).

As it emerges from the Anglo-American approach to educational testing, the WISC is a very Western European- centred test, and as such it is prone to cultural bias. For example, Similarities (administered in English) may not be appropriate for a child who is a non-native English speaker. Although the WISC has been translated in six languages, it remains situated within a Western European understanding of intelligence, test-wiseness and formal taxonomic education, that might not be valid when used outside of those cultural parameters.

1.4.3 Summary of critiques of existing Concept Formation measures for Young Children

Eight different tests were identified as existing measures of concept formation for six- and seven-year-old children. The WISC Similarities and Matrix Reasoning, together with Ravens SPM represent the measures that are most established. In the case of SPM and Matrix Reasoning children may use different problem-solving strategies. In the case of Similarities, there are important considerations of cultural bias embedded from the overall design to the item level.

So, the review of the existing tests suggests that they are undermined by the number of confounding variables, including motor perseverance, high inhibition and working memory demands, need for verbal responses, and the format of the test are all factors that affect the validity and reliability of the scores. Cultural biases add an additional layer of confounding variables that affect performance and therefore the validity and reliability of scores. The interaction between scoring and these confounding variables is not well understood or explored in the tests reviewed above.

1.5 The Alien Game

Pavitt (2017) conducted the first pilot test of the Alien Game, which is a novel game-based test of concept formation for children. It was designed on the basis of a measure by Delis et al. (2001), called the D-KEFS 20-Question Task, by adapting it into a game-based task by using the board game *Guess Who* (® Hasbro Games) as a template.

The D-KEFS 20 Questions Task can be used to test concept formation in children and adults, covering an age range from 8 years to 89 years. In this task, the person being assessed has to identify one target item from among 24 objects presented in a visual array. The pictured items belong to two main taxonomies of living and non-living things, which are further subdivided into smaller subcategories (e.g. animals or plants; tools or furniture). The person completing the task can ask up to 20 closed yes or no questions (e.g. “Does it have wheels?”) per trial to identify the target. The best strategy to solve the task is to ask questions that, regardless of outcome, will eliminate as many non-target pictures as possible. For example, as there is an equal number of living and non-living items, the best initial question would be to ask if the target item is a living thing, as this would eliminate half of the items. The identification of this strategy requires concept formation, to deduce the hierarchy of taxonomies or rules into which the items can be grouped, and to address the highest order grouping available each time, in order to eliminate as many pictures as possible.

The board game *Guess Who* (® Hasbro Games) is based on a similar strategy, using 24 human characters instead of groupings of hierarchical taxonomies. The grouping rules by which characters can be eliminated rely on the physical features and accessories of each character, e.g. whether

they wear glasses, or have freckles. Pavitt (2017) utilised this game-based design since it would possess greater potential as an assessment tool specifically designed for children, as opposed to being a downward adaptation of an assessment originally designed for adults. In order to negate cultural biases, the stimuli of the Alien Game are intended to be free from any specific cultural references. In addition, the test is designed to mitigate any direct resemblance to similar commercially available games, as this would introduce a direct confounding variable of familiarity with the stimuli. The game-like format adopted is intended to measure a child's response to what could be a novel age-appropriate activity (Helms-Lorenz & Van De Vijver, 1995) as opposed to a learnt response or strategy.

In the original Alien Game, the child being assessed can ask up to twenty closed (yes or no) questions to identify a target alien on a board of 24 aliens (the original array is given in Appendix 6.1). The aliens have features from among 13 attributes, provided in Table 1 in Section 2.2.2. Some attributes may or may not be present (dichotomous attributes, such as the alien having wings or not) and some can have three possible variations (trichotomous attributes, such having either as a triangular, rectangular, or circular body shape). Pavitt (2017) assigned features to the original aliens in the same proportions to the *Guess Who* game, meaning the allocation was not equal (for example, the number of aliens with wings was not the same as the number of aliens without wings).

The rationale for the design and variable allocation of attributes by Pavitt (2017) was to increase the game-like quality of the task by introducing the element of chance, and therefore enhance its child-friendly nature. However, this design also introduces the element of luck, that is a high score in the

game could be the result of an early lucky guess as opposed to a strategy based on induction and concept formation abilities.

To mitigate the effect of luck, one of the scoring systems piloted by Pavitt (2017) was an adaptation of the D-KEFS (Delis et al., 2001) scoring, called the Weighted Achievement Score (WAS). This system awards most points if the person completing the task asks the optimum number of questions to identify the target alien. The optimum number of questions is calculated based on an error-free strategy that would lead the examinee to ask questions with the highest *minimum elimination potential* (MEP; the MEP is the number of aliens in the smallest of two groups defined on the basis of the attribute in the participant's question). These are low risk questions, because they tend to divide the (remaining) aliens in equal groups, so that if the guess is wrong a large number of aliens would still be eliminated. The optimum strategy is to ask questions with a minimum elimination potential as close to 50% as possible.

In Pavitt (2017) the optimum number of questions was five, yielding a WAS score of five. If more questions were asked, this indicated a weaker strategy or more errors (i.e. questions that did not eliminate due to repetition or wrong format) and there was a point subtracted from the score as penalty for each additional question. If the alien was identified in fewer than five questions, this indicated a 'lucky guess' rather than conceptual strategy. Points increased the closer the participant got to asking five questions, with only one point being awarded if the participant asked one or two questions, two points for three questions, and four points for four questions.

In addition to WAS, Pavitt (2017) used the Initial Abstraction Score (IAS), the Abstraction Score (AS), and the Learning Slope (LS). The IAS is the

Minimum Elimination Potential of the first question. The AS is the Minimum Elimination Potential total over the first three questions asked. As there are four trials in total, the sums of WAS, IAS and AS are calculated to obtain their totals. Finally, the LS was the difference of the minimum elimination potentials of the first versus last trials, which was a measure of improvement over the trials in the task.

Pavitt (2017) piloted the Alien Game with children aged seven to eleven years. One aim was to evaluate which of the above scoring metrics was more suitable as a concept formation score. A second aim was to analyse the strategies of participants to determine whether they were using similar approaches to the game, i.e. that the Alien Game was consistently measuring participants' efficiency in perceiving the aliens' characteristics, forming taxonomies of the aliens, and asking questions to eliminate as many of the non-target aliens as quickly as possible. A content analysis of the participants' questions showed that participants were consistently asking elimination questions in the correct format on the basis of the aliens' shared attributes. Pavitt (2017) therefore concluded that the Alien Game had good potential.

Pavitt (2017) found that her participants of all ages used similar strategies when playing the Alien Game. They tended to ask questions that were 'constraint seeking' rather than 'hypothesis testing'.

However, it is not clear whether the game parameters and associated scoring system were effective in translating performance into reliable scores. The Initial Abstraction Scores (IAS) and Abstraction Scores (AS) were moderately correlated (r=.331). The highest correlation was found between

IAS and WAS (r=.390). The correlation between AS and WAS was low (r=.146).

There was a moderate negative correlation between IAS and age (r=-.515, p=.029), suggesting that older participants performed less well than younger participants on this measure. The coefficient sizes indicate that these scoring systems may reflect on disparate aspects of performance which might rely on different abilities.

In her study, Pavitt (2017) used two sub-tests from the WISC to measure participants' concept formation: Similarities (verbal reasoning), and Matrix Reasoning (non-verbal abstract thinking). Both of these are relevant for concept formation. In relation to Similarities, the IAS and WAS yielded positive correlations. In relation to Matrix Reasoning IAS and WAS yielded negative correlations. The meaning of this discrepancy was not clear. Pavitt (2017) also reported that the Alien Game was rated higher in 'enjoyability' than Similarities and Matrix Reasoning.

The findings of Pavitt (2017) have two main implications. The first is that children seem to be using similar strategies when playing the Alien Game, therefore providing evidence that it lends itself as an instrument to measure concept formation. Secondly, the statistical trends between the Alien Game and WISC Similarities indicated that there may be some shared variance in what they measure, i.e. concept formation.

1.6 Summary, study aims and research questions

Concept formation is an important aspect of the executive functions, and it refers to the ability of a person to categorise objects, according to their features and connections, which are represented as concept. Objects can be categorised in many different ways, therefore a successful application of concept formation to a given situation is to determine which concepts are most pertinent to be used as grouping criteria. This requires the use of inhibition and working memory, which are separate executive functions that assist in concept formation. Inhibition is essential in supporting the flexible use of new concepts, while working memory is essential for holding in awareness the different concepts and strategies to be applied.

Overall, concept formation is an ability that merits attention from a developmental and educational perspective, as it supports learning and adaptive functioning. It is therefore important for difficulties in concept formation to be detected as early as possible. There are several concept formation tests available for adults, some of which have been extended downwards to be used with children. Downward extension is not the most appropriate design for cognitive assessments in children, as there is a risk of reduced construct and ecological validity. In contrast, there are few concept formation tests available for young children below eight years of age.

The Alien Game is a game-based concept formation test designed specifically for children, and is designed to mitigate cultural bias. It has been piloted with children aged seven to eleven years, and has shown potential to be a consistent measure of concept formation. The present study is a pilot of

the Alien Game for a younger age group, specifically aimed at six and seven year old participants, to evaluate its feasibility as a concept formation measure for that age group.

For the purposes of the present study, the following considerations were made to design the Alien Game pilot for a slightly younger age group. Firstly, the Alien Game is a predominantly visual-verbal task, and therefore eliminates the need for motor responses, which may be prone to motor perseverance. The need for verbal aspects is limited to understanding the instructions and being able to express yes/no questions, but it does not require taxonomic knowledge, or other crystalised knowledge to complete the task. This feature of the Alien Game, i.e. that it is a measure of ability and not acquired knowledge, as well as the relatively culturally neutral stimuli used, help to reduce the cultural biases of the measure.

Inhibition and working memory abilities are shown to undergo maturational changes around the age of seven years. For this reason, a format variation of the Alien Game was used in the present study, to mitigate potential demands upon inhibition and working memory. The novel format involved presenting the aliens on individual cards that could be flipped to a blank obverse side over if the alien was eliminated. This had the potential to reduce errors that could be caused by difficulties with inhibition, e.g. repeating questions elicited by the other features of the non-target stimuli. In considering the role of working memory, there was potential to reduce the demands on short-term stores and therefore help in the central executive conceptual processing of the alien's features. Card flipping would be done by the person administering the test in the interest of eliminating the need for manual handling of the stimuli by the participants.

A further change from the original pilot was the inclusion of a teacher reported objective measure of real-world executive function for the participants, which would help to address the role of inhibition and working memory in playing the Alien Game.

Finally, given the age of the participants, and to reduce the potential of frustration and fatigue, the number of questions permitted per trial was reduced to 10. This shortened the length of the participation sessions, reducing the time away from the classroom, and minimising stress and fatigue.

The aims of the present study were to pilot the Alien Game with six and seven year olds. Two conditions were piloted, Board and Cards, to evaluate which may be more appropriate at measuring Concept formation in this age group.

The research questions were:

- Will there be a difference between the Boards and Cards conditions of the Alien Game with respect to performance scores, number of questions asked, and number of errors?
- How do the Alien Game scores of six and seven year olds correlate with existing measures of WISC-IV Similarities and Matrix Reasoning?
- How do the Alien Game scores of six and seven year olds correlate with objective measures of executive function reported by a teacher?
- Finally, if the participant sample would allow a General Linear Model analysis with respect to influence of cultural factors, what is the influence of cultural factors, such as languages spoken at home, on the performance in the Alien Game?

2. METHODS

2.1 Epistemological Position

Research on executive functions and concept formation can be seen as part of wider cognitive science, which is heavily influenced by realism and positivism.

The epistemological position chosen for this study is that of critical realism (Bhaskar, 1998). Critical realism can provide the framework for a critical dialogue that can examine the methodology, results and conclusions of cognitive research with the necessary self-awareness that the researcher is not an asocial being (Cruikshank, 2003). As discussed in the previous section, the belief that the researcher can interpret results to theorise on a directly accessible and objective truth has historically led to epistemic oppression (Buchanan et al., 2021). Critical realism does not, though, make the post-modernist and constructionist claims that cognitive sciences and terminology within that body of research are ephemeral manifestations that do not transcend the researcher's discourses within their limited socio-historical context (Cruishank, 2003).

Critical realism sits between these two epistemological positions, and therefore can provide theoretical balance in a subject matter that touches on the biological as well as the socio-cultural aspects of developmental cognitive psychological processes. Critical realism accepts that an individual

has enough free will to exert agency within their self and their environment. This sense of self may be influenced by the person's social context, but is not determined by it (Cruishank, 2003). This has implications for both the researcher and the participants of a study.

The researcher can seek to understand a reality that exists separately to any representation of it that the researcher will ever attain Still, the researcher can create models of that reality to approximate it. This position accepts that the researcher's knowledge will be limited, and any models or theories created will be fallible (Cruishank, 2003). Models and theories are based on data derived from participants who are influenced by their own agency as well as the structures within their wider sociocultural context (Cruishank, 2003). These structures may include the participants' education, the way an educational system may have interpreted their abilities, whether they have received support, whether they have been subjected to discrimination and other forms of social oppression, the way their parents' socioeconomic status has affected their learning, and so on.

A critical realist theory aims to understand how agency and structures are linked, but it does not afford a theory that would fully describe that relationship (as this would essentially be a positivist position). Instead, it seeks to create a meta-theoretical framework that can guide our understanding of the particular relationships that may exist between types of agency and types of structure (Cruishank, 2003). Within this position, variables, and by extension relationships between variables, do not have a factual relevance. Instead they are seen a conceptual interpretation of an ultimately inaccessible objective reality (Cruishank, 2003).

With regards to this present study, the implications of a critical realist position are that executive functions, including concept formation, are terms that represent models of cognition that operate within the self and may be seen as part of a person's agency. The structures that are seen as influential on executive function and concept formation are the measurement tools used, which may be more or less prone to confounding variables and testing biases, as discussed above.

2.2 Methodology

2.2.1 Study Design

This is a cross-sectional study to explore whether the Alien Game can be extended to six year old children, an age group which has not yet been assessed by the Alien Game. Based on Pavitt (2017), it was found that children of ages of seven and eleven years use consistent strategies to solve the Alien Game. This current study, therefore, has included seven year olds as well as six year old children to enable between-participant score comparisons. Within-participant comparisons will be made to evaluate if there is a more developmentally appropriate format of the Alien Game. In addition, parts of this study have a correlational design to examine the relationship between the scores of the Alien Game and established concept formation measures including the WISC (WISC-IV Similarities and Matrix Reasoning; Wechsler, 2003) and a teacher-reported measure of executive function, the CHEXI (Child executive function Inventory, Thorell & Nyberg,

2008), which may indicate the Alien Game's concurrent and predictive validity respectively.

The WISC-IV Similarities and Matrix Reasoning subtests were used to address the concurrent validity of the Alien Game, in terms of its ability to measure concept formation. The WISC-IV was used by Pavitt (2017), and therefore was chosen for comparison purposes. The WISC-IV is an established cognitive measure and its use is appropriate in a research context and for the exploration of construct validity, i.e. to explore whether there is a correlation between the scores of the Alien Game and scores of Similarities and Matrix Reasoning. The WISC-IV has high ecological and criterion validity for clinical as well as non-clinical populations and has been used in research widely (Yeates & Donders, 2005).

The CHEXI was used as a measure of executive function, including working memory, planning ability, inhibition and regulation. It is a freely available questionnaire, making it accessible for the present study, and it is reported to have high ecological validity and construct validity (Thorell & Nyberg, 2008). However, its working memory and inhibition scores have been found to be more robust than its planning and regulation scores (Thorell & Nyberg, 2008). The CHEXI questionnaires were completed by the teacher for each of the participants.

Finally, a General Linear Model Analysis may be utilised if data meet normality thresholds, to evaluate the influence of cultural factors (e.g. language spoken at home) on concept formation scores.

2.2.2. Alien Game Material Development

The Alien Game set (Figure 3) was created using Version 27 of Adobe Illustrator (Adobe Inc., 2022). The set contained a total of 24 aliens presented in a 6x4 grid, each alien occupying a 6.5cm x 6.5cm tile to make them clearly visible to participants. The aliens were based on the original set by Pavitt (2017, see Appendix 6.1). In Pavitt (2017), the aliens had a total of 13 attributes, such as number of eyes, legs, antennae (shown in Table 1). The original colours used were grey, blue and yellow, as these are accessible to people with colour blindness. The number and allocation of attributes was based on the properties of the *Guess Who* (® Hasbro Gaming) board game, and were unequally distributed among the stimuli. This allocation increased the likelihood of lucky guesses as a chance factor while participants were playing the game. To reduce the effect of luck in the present study, an even allocation of attributes was used (see Table 1). For example, in Pavitt (2017) there were six aliens with eyebrows and 18 without. Therefore, the way a participant formulated a yes/no question ("Does it have eyebrows?" versus "Does it have no eyebrows?") could have resulted in the elimination of one third vs two thirds of the aliens. In the updated version, there are 12 aliens with eyebrows and 12 without, therefore whichever way the question would be formulated, it would result in the same elimination of aliens.

In elimination assessment tools (such as the 20 Questions Task of the D-KEFS, Delis et al., 2001), the optimum strategy is to ask an elimination question on an attribute shared by 50% of the stimuli, since, regardless of the outcome, half the stimuli will be eliminated. In the Alien Game, given the equal distribution of attributes, it is possible to ask two to three consecutive

questions, eliminating 50% of the stimuli with each question (i.e. 12 aliens out of 24, and then six aliens out of 12, and then three aliens out of six). When there are three aliens left, it is not possible to eliminate 50%. Therefore, depending on an element of luck, a further one or two questions may be needed to identify the target alien. As a result, the optimum strategy would consist of asking yes/no questions, based on bi-variate attributes, that could range between four to five questions.

Two different sets of aliens were created, both of the same overall size. The first set was a grid of aliens printed on an A3 sized paper and laminated. The second set were individual aliens on individual cards that were arranged in the same array as the grid. The use of individual cards meant that aliens could be flipped over as the participants were eliminating them, so that non-targets could be physically removed from sight. The grid set would involve more taxing use of working memory, as the participants had to remember which questions had been asked and which aliens had been eliminated. The card set was not so demanding, as only the non-eliminated aliens were visible and the participants could focus on identifying relevant attributes that they could use to identify the target.

Figure 3.

The modified Alien Game set.

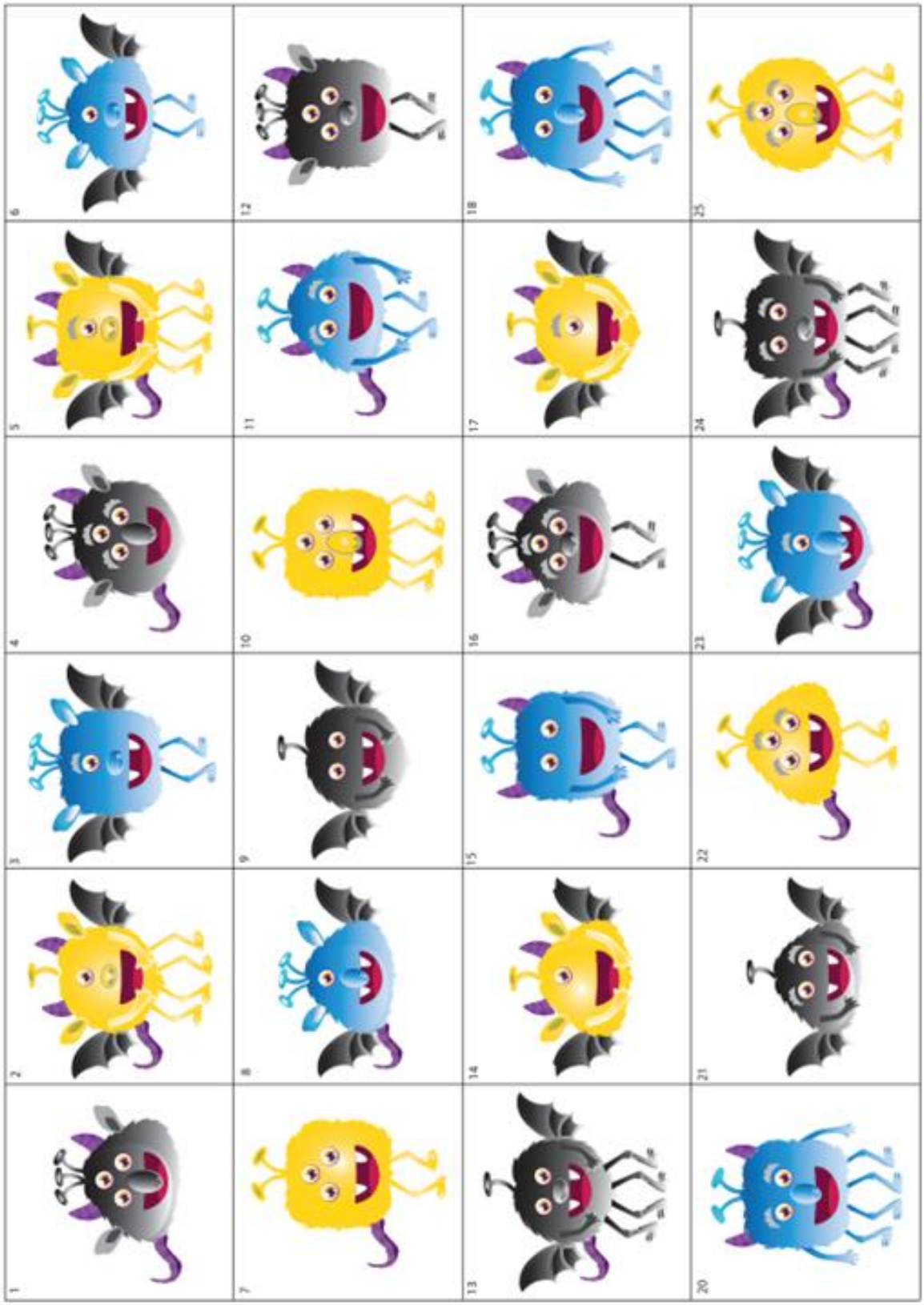


Table 1.

Original allocation of Alien Game attributes in Pavitt (2017) and in the updated set.

Attribute	Variations	Updated Set	Pavitt (2017) Original set
		N	N
Horns	Yes	12	16
	No	12	8
Tail	Yes	12	12
	No	12	12
Ears	Yes	12	19
	No	12	5
Arms	Yes	12	4
	No	12	20
Wings	Yes	12	11
	No	12	13
Eyebrows	Yes	12	6
	No	12	18
Teeth	Yes	12	7
	No	12	17
Shape	Circle	8	10
	Triangle	8	7
	Square	8	7
Colour	Grey	8	8
	Blue	8	9
	Yellow	8	7
Number of Antennae	One	8	13 (no antennae)
	Two	8	5 (one antenna)
	Three	8	6 (two antennae)
Number of Eyes	One	8	6
	Two	8	12
	Three	8	6
Nose	No nose	8	8
	Small nose	8	11
	Big nose	8	5 (trunk)
Number of legs	None	8	12
	Two	8	8
	Four	8	4

2.2.3 Materials Used

The following materials and equipment were used for the study.

- Alien Game set
- WISC-IV Matrix Reasoning and Similarities subtests (stimuli, scoring and interpretation resources).
- CHEXI questionnaires
- Study information sheets (for parents and pupils)
- Consent forms (for parents and pupils)
- Child debriefing forms
- Standardised instructions
- Table, chairs, pen and paper

2.2.4 Procedure

The testing procedure was the same for all participants. Testing sessions took place in a quiet room of the participants' school during standard school hours. At the start of each testing day, the researcher provided a brief description of the task in the classroom to help the children decide whether they wanted to participate (provided their parents had not opted out). All children expressed the wish to participate. Participation was in the order names appeared on the school attendance register.

For their individual testing session, participants were seated at a desk. They were given a copy of the Information sheet, which was explained and the opportunity to ask questions was provided. Once participants gave their assent to continue with the study, the researcher asked a set of demographic

questions. These included age, date of birth, primary language spoken at home, languages spoken by the participant, and parents' occupation. The child's gender was also recorded.

Cognitive tasks were then completed in the following order: Alien Game, Similarities, and Matrix Reasoning. Once these tasks were completed, participants were given a rating question (given in Appendix 6.2) that asked them how enjoyable they found the Alien Game (after the seventh participant, this question was altered to include ratings of the Similarities and Matrix Reasoning Task). Finally, children were given another opportunity to ask any questions about the study, were given the debriefing sheet and were reminded that their parents can contact the researcher if they had any questions, or if they wished for their child's data to be withdrawn.

The instructions for the Alien Game, Similarities and Matrix Reasoning were presented in a standardised manner (see Appendix 6.3 for the instructions of the Alien Game). The Alien Game involved four trials. Each version of the set was used twice in a counterbalanced order (for the counterbalancing schedule, see Appendix 6.4). Responses were recorded verbatim by the researcher on the record forms.

The administration of the Alien Game involved the researcher reading the standardised instructions (Appendix 6.3) explaining the aim of the game, i.e. that the participants had to guess the target alien as quickly as they could and within a maximum of 10 questions that could be answered with a "yes" or a "no". If the participants asked questions in the wrong format the researcher answered that the participant needed to think of a "yes or no" question to ask. If an irrelevant or redundant question (e.g. "Does it like reading?" or "Does it have wings?" when all remaining aliens had wings) was

asked, the researcher answered that they could not answer it and directed the children's attention back to the Alien Game stimuli, and asked them to think of a different question. For the trials where the cards were used, the researcher flipped over any cards that had been eliminated by the participants' question. This was to mitigate any motor disinhibition as per Greve et al. (2000). For example, it made it less likely that a participant would ask a question based on the attributes of the aliens around the alien that had been eliminated, simply because their hand had gone in that direction and their attention was focused on the nearby cards.

2.2.5 Data Collection and Analysis

The outcome variables of the study were: the participants' demographics, their scores in the Alien Game, Similarities and Matrix Reasoning tasks, their CHEXI scores, and their ratings of enjoyability of the tasks.

Demographic variables collected were: the child's age at testing, gender, languages spoken at home, main language spoken at home, and parental occupation.

The raw and age-scaled scores of Similarities and Matrix Reasoning were included in the analysis. The raw scores of the CHEXI were used, as it has not been normed for children aged six and seven years.

There were a number of Alien Game scores in the present study. Table 2 explains how these variables were calculated. They were based on the adopted D-KEFS (Delis et al., 2001) scores and scoring method, which were used by Pavitt (2017).

The Abstraction score (AS), Initial Abstraction score (IAS), and the Learning Slope (LS) are based on the Minimum Elimination Potential (MEP) of the questions asked. The IAS is the MEP of the initial question, the AS is the sum of MEP of the first three questions asked in a trial, and the LS is the difference of the AS of trial 4 minus the AS of trial 1.

The Weighted Achievement score (WAS) is calculated based on the number of questions asked. The same scoring was used as Pavitt (2017) (see Table 2). This means that the Total Number of Questions (TNQs) was calculated to determine the WAS.

Similar to Pavitt (2017), and D-KEFS (Delis et al., 2001) the questions participants asked were coded. The terms used in the present study were Conceptual Questions (CQs) to describe questions that eliminated more than one aliens. Pseudo-Conceptual Questions (PCQs) were the questions that had an appropriate yes or no format, but eliminated only a single alien (e.g. “Is it number 8?”).

There were two error question types. Redundant Questions (RedQs) included all questions that did not eliminate any aliens. This could be because a question referred to an attribute that none of the remaining aliens had, or because it referred to properties of the aliens that were not part of the game (e.g. “Does it like to read books?”). Any questions asked in the wrong format were included in this category as they failed to eliminate any aliens. The second e the number of repeated questions. Repetitive Questions (RepQs) was the second error question type, and it included any questions that had already been asked in the current trial.

It will be part of the data analysis to explore the suitability of these scores as assessment indices. This will be determined based on the variables' distribution range and concurrent validity with WISC-IV Similarities and Matrix Reasoning.

Table 2.

The variables derived from participants' performance in the Alien Game

Variable	Scoring
Abstraction Score	The total abstraction score is calculated by adding the minimum elimination potential of the first three questions of each trial, summed across the four trials. The minimum elimination potential is the lowest number of aliens that can be eliminated using a yes/no question, regardless of the outcome.
Initial Abstraction Score	The Initial Abstraction score is calculated by adding the minimum elimination potentials of the first question of each trial, summed across the four trials.
Weighted Achievement Score	This is a conversion of the total number of questions asked to a score that represents how close that was to the number of questions expected if the optimum strategy were followed. 5 questions= 5points; 4 or 6 questions= 4 points; 7 questions= 3 points; 3 or 8 questions= 2 points; 1,2,9 or 10 questions= 1 point.
Learning Slope	The Learning Slope is equal to each participant's AS score in Trial 4 minus Trial 1.
Total Number of Questions	This is the number (up to 10) of questions participants could ask to find the target alien
Conceptual Questions	Total number of questions that eliminated more than one aliens.
Pseudo-Conceptual Questions	The total number of questions that had an appropriate Yes or No format, but eliminated only a single Alien (e.g. "Is it number 8?").
Redundant Questions	This is the number of questions that did not eliminate any aliens, due to referring to an attribute that none of the remaining aliens have, or because it refers to properties of the aliens that were not part of the game (e.g. "Does it like to read books?"), or because the questions was in the wrong format.
Repetitive Questions	This was the number of questions that had already been asked in the current Trial.

2.2.6 Participants

The recruitment goal for the present exploratory study was at least 36 participants, i.e. approximately 18 six-year-olds and 18 seven-year-olds. This was deemed both achievable within the logistical context of the research project, and was twice the size of the Pavitt (2017) sample. Therefore, the aim was that each year group included in the present study would be of the same size as the total sample of Pavitt (2017). This would allow for clearer comparisons between the performance of six-year-old and seven-year-old participant groups to each other, and comparisons between performance in the present study and Pavitt (2017).

Primary schools in London and the South of England were contacted for recruitment purposes. Sixty-one primary schools were invited to participate in the study. The vast majority of these were based in the wider London area, and a small number were in cities in the South East and South West of England. Only one school, based in London, agreed to collaborate. The remaining schools did not engage in requests for a call back from their SENCo or other members of the school management team. Only a few responded to emails expressing their inability to engage with the research due to logistical reasons, such as not having the necessary number of staff to support the researcher or not having sufficient space.

The school in London that agreed to collaborate, facilitated recruitment in their Year 2 group, ultimately leading to 14 students being recruited. Participating children therefore formed a convenience sample, as they were not randomly selected. Children were included in the study if they were within the age range (six years to seven years) and their parents had not opted out of the study (for more details, see Ethics section below). Participant

exclusion criteria from this study were if parents declined consent and/or if the child had behavioural or communication needs that would interfere with their ability to assent or participate. No children, from those whose parents consented, were excluded from participating in the study.

Five participants did not complete all four trials of the Alien Game, three participants did not complete a fourth trial, and two participants did not complete a third nor fourth trials (see Appendix 6.4). As a result, there were no available scores for a second trial of the Cards condition for four participants in total; and no available scores for a second trial of the Board condition for two participants, one of who did not have any scores for the first trial of the Board condition. Accordingly, it was preferrable to focus some of the analysis on the first trials of the two conditions. This means that 13 participants were included in the final sample for analysis, and the one participant who completed none of the Board condition trials was removed. A suspected reason for non-completion is thought to be that the researcher checked between trials if the participants would like to continue with the Alien Game, which gave them more opportunities to decline.

There were seven 6-year olds, with a mean age (in months) of 80.2 (SD=3.2, minimum age 74.9 months, maximum age 83.6 months) and six 7-year olds with a mean age of 87.2 months (SD=2.1, minimum age 84.1 and maximum age 89.9). The mean age (in months) for the entire sample was 83.4, (SD= 4.5).

There were 7 male and 6 female participants. The range of reported languages spoken at home was between one and three. Four participants spoke one language at home, six participants spoke two languages, and four participants spoke three languages. With respect to the main language

spoken at home, ten participants reported speaking English as the primary language spoken at home, two could not differentiate between English and other languages, and one participant reported that English was not the main language spoken at home.

All participants were asked about their parents' professions, but most answers provided were not clear. Therefore this information will not be considered further as a variable.

Descriptive scores of participants' scores on the Similarities and Matrix Reasoning Subtests of the WISC-IV, and the teacher's CHEXI questionnaires, are presented in Table 3.

Similarities Scaled score is very close to the average (mean = 10, SD = 3) There was a wide range of scores, with a minimum of 2 and a maximum of 17. The Matrix Reasoning Scaled scores were slightly below expected. The range was between a minimum of 5 and a maximum score of 12. This indicates that the sample was representative of the average population with respect to their Similarities scores but somewhat below average in terms of their Matrix Reasoning scores.

As some data appeared non-normal from a visual inspection of Skewness and Kurtosis values, and due to the small sample size, the non-parametric Mann Whitney U-test was used for group comparisons around the Similarities and Matrix Reasoning Scaled scores and CHEXI Working Memory and Inhibition raw scores. The Exact Tests resampling procedure available on SPSS was used to yield more robust results. Comparisons were based on age (six years versus seven years), gender (male versus female), and primary language spoken at home (English only primary language

spoken at home versus English not the primary or only primary language spoken at home).

As shown on Table 4, there was only one significant difference found in the group comparisons. This was the Matrix Reasoning scores with respect to age, with six year olds scoring higher than the seven year olds. All other score differences were non-significant.

Table 3.

Descriptive statistics for WISC-IV Similarities and Matrix Reasoning and CHEXI scores.

Variable	Mean	SD	Min	Max	Skewness (SE=.616)	Kurtosis (SE=1.19)
Similarities Raw Score	11.31	7.465	0	23	-.148	-1.086
Similarities Scaled Score	10.15	4.947	2	17	-.382	-1.029
Matrix Reasoning Raw Score	11.92	2.660	7	17	.097	.146
Matrix Reasoning Scaled Score	8.15	2.154	5	12	.712	-.217
CHEXI WM Raw Score	20.77	5.732	9	20.77	.251	1.529
CHEXI Planning Raw Score	9.15	2.230	4	9.15	-.494	1.699
CHEXI regulation Raw Score	12.08	2.691	5	12.08	-1.369	3.576
CHEXI Inhibition Raw Score	14.38	2.256	11	14.38	-.002	-1.389
CHEXI Total Raw Score	56.38	11.758	29	56.38	-.560	1.849

Table 4

Group comparisons of Similarities, Matrix Reasoning Scaled scores, CHEXI Working Memory and Inhibition based on gender, age, and primary language spoken at home.

							Mann-Whitney U-Test		
			Mean	SD	Min.	Max.	U	Z	Exact Sig.
Gender	Sim.	Female	9	5.2	2	15	15.5	-.79	.469
		Male	11.1	4.9	3	17			
	MR	Female	7.5	1.9	5	10	15.5	-.81	.461
		Male	8.7	2.4	7	12			
	WM	Female	22.3	5.2	17	32	14	-1.0	.346
		Male	19.4	6.2	9	30			
	Inh.	Female	14.7	2	12	17	17.5	-.51	.644
		Male	14.1	2.6	11	18			
Age	Sim.	6yo	11.4	6.0	2	17	11	-1.4	.166
		7yo	8.7	3.3	3	12			
	MR	6yo	9.3	2.4	5	12	6	-2.2	**.023**
		7yo	6.8	0.4	6	7			
	WM	6yo	19.4	6.9	9	32	11.5	-1.4	.192
		7yo	22.3	3.9	19	30			
	Inh.	6yo	13.9	2.2	11	16	14	-1.0	.347
		7yo	15.0	2.4	12	18			
Lang.	Sim.	English	9.3	5.0	2	17	8.5	-1.1	.297
		Not Engl	13.0	4.4	8	16			
	MR	English	7.9	2.1	5	12	10.5	-.79	.524
		Not Engl	9.0	2.6	7	12			
	WM	English	21.3	6.4	9	32	9.0	-1.0	.350
		Not Engl	19.0	2.6	17	22			
	Inh.	English	14.6	2.2	11	18	12.0	-.52	.664
		Not Engl	13.7	2.9	12	17			

Note. N=6 for females, N=7 for males. N=7 for 6-year olds, N=6 for 7-year olds. N=10 for English as main language, N=3 for English not the main language. Significant p values are in **bold**. For the Mann-Whitney U-Test, the Exact Tests resampling procedure available on SPSS was used to yield more robust results.

2.2.7 Ethics

The present study was granted Ethical Approval by the Psychology Department's Research Ethics Committee of the University of East London (please see Appendix 6.5).

The first step in recruiting participants was to establish collaborations with schools in the wider London area. One school agreed to collaborate. A choice between an Opt-In and Opt-Out consent process was given to the school, and they chose the Opt-Out process. This involved the head teacher signing the *loco parentis* consent form (Appendix 6.5). Then the school emailed all parents the Study Information Sheet (Appendix 6.5) and sent an Opt-Out parental consent form (Appendix 6.5). A period of two weeks was given to the parents to consider the information and contact the researcher with any questions. After that period, the researcher visited the school for data collection. Children were invited to participate only if their parents had not opted out.

The Child version of the Participants' Information Sheet and the Consent Form (Appendix 6.5) was presented and explained to all participants individually at the start of each study session. Given the age of the participants, simple language used to explain the Information sheet, which was repeated as much as necessary, and ample opportunities were given to all participants to ask questions. If they agreed to continue with the study based on the information provided, they were asked to sign the consent forms. At the end of the study session, participants were invited to ask any questions again, and a copy of the debriefing form (Appendix 6.5) was given to them. The contact details of the researcher and their supervisor were available on all forms for the parents to communicate any questions.

Parents and participants were informed of their right to withdraw from the study at any point during the testing session, and that afterwards their data could be withdrawn within a given deadline (two weeks post data collection) by contacting the researcher. The study was risk assessed (for the Risk Assessment, see Appendix 6.5) and was deemed to be low risk to children. There were no deception elements in the study and no risky or hazardous procedures. As per the risk assessment, in the event of a participant showing signs of distress the study would be paused, assistance sought from the teacher or school staff, and testing would only resume if the child expressed the wish to continue. No children showed signs of distress during their participation.

Confidentiality of personal and identifiable information was protected via the following methods. The name of participants were used only on the child's consent form, which were scanned and electronic copies saved on secure online storage provided by University of East London (OneDrive facility). Codes were used on all other forms, including the record forms and CHEXI questionnaires. These forms were scanned and securely stored separately on the secure University OneDrive. An electronic file was created and saved on the OneDrive to keep a record of code allocation to participants in the event that a data withdrawal request was made. Once scanned and securely stored in electronic form, all paper copies were destroyed via confidential waste. The data analysis files that were created contained only anonymised data (i.e. participants' given codes) and were kept securely on the OneDrive. Participants and their families were informed that anonymised study findings may be disseminated via research publications. Only anonymised data files will be retained beyond the completion of this project and will be deposited in

the Research Repository of the University of East London, under the care of the researcher's Dissertation Supervisor. All other electronic files on the researcher's personal OneDrive account will be safely deleted.

3. RESULTS

3.1 Approaches of Analysis

The analysis of the data collected was performed on SPSS Version 28 (IBM Corp., 2021). Initially, an exploratory data analysis (EDA) was performed to identify any outliers, data entry errors and data distributions. Due to the sample size, the EDA was mainly based on the visual inspection of histograms and box plots. In addition, skewness (a value of less than 1), kurtosis (a value of less than 3) and the Shapiro-Wilk test of normality were calculated (please see Table 10 in Appendix 6.6) The Shapiro-Wilk Test was preferred as it has better accuracy than other normality tests when the participant sample is small (Razali & Wah, 2011).

As explained in the previous section, due to missing data the analysis of the Alien Game scores was focused on scores based on Trial 1 of the Cards condition and Trial 1 of the Board condition.

As it is shown below, the variables involved in the analysis were predominantly non-normal. For this reason, and due to the small size of the sample, non-parametric tests were used for subsequent analyses. To evaluate the relationship between variables, Spearman's Rank correlations were used. Coefficient values were interpreted as effect sizes according to Cohen (1988): .10 to .29 were interpreted as small; .30 to .49 as moderate and 0.50 and over as large. The Mann-Whitney U Test was used for between

participants comparisons and the Wilcoxon Signed Rank Test was used for within participants contrasts. Where possible, the Exact Tests resampling procedure available on SPSS was used to yield more robust results.

3.2 EDA of Alien Game Variables

The Alien Game scores were the IAS, AS and WAS. In addition the types of questions participants asked were explored. These were the Total Number of Questions (TNQs), the number of Conceptual Questions (CQs, meaning that they were of an appropriate Yes/No format and eliminated more than just one alien), the number of Pseudo-Conceptual Questions (PCQs, meaning that they were of an appropriate Yes/No format but targeted only one alien at a time), and number of Redundant (RedQs, meaning that they did not eliminate any aliens) and Repetitive Questions (RepQs, meaning that they were in an appropriate Yes/No format, but they had asked that question already).

A combination of visual inspection of histograms and box-plots and a consideration of The Shapiro-Wilk *p* value (please see Table 10 in Appendix 6.6) indicated that most Alien Game variables were not normally distributed. However, the Board AS, Board CQs and Card TNQs were indicted to be normally distributed. The remaining variables appeared to have a combination of negative skewness, bi-modality and discontinuous data.

3.3 Consideration of Applicability of the Alien Game

From an applicability perspective, it is important that the Alien Game scores for six and seven year old children show sufficient variability and an absence of floor and ceiling effects. These qualities mean that a score has the potential to capture a sufficient range of concept formation ability and that it could lend itself for standardisation as part of future development of the measure.

In Pavitt (2017) IAS, AS and WAS of Trials 1-4 would be summed, whereas in the present study they are based on a single Trial. However, from the available data, it appears that none of the IAS, AS or WAS yielded a floor or ceiling effect. As shown on Table 5, there was consistency in terms of the range of score values between the Cards and Board conditions. The mean scores fell relatively in the middle of the range of values. In addition, the ranges of values obtained by participants matched the possible range of values. These are 0-12 for the IAS, as the highest MEP of the first question can be 12; 0 to 21 for the AS, as this is the total MEP of aliens that can be eliminated through the first three questions; and 0-5 is the range of scores for the WAS). Out of the three scores, AS had the widest range of participants' scores.

With respect to the variables based on types of questions asked, the TNQs had a similar range to the IAS, and the CQs, RedQs and RepQs had as a narrow range as the WAS. In addition, they appeared to have a ceiling effect, such as the TNQs, especially in the Board condition; or a floor effect, such as

the RedQs and RepQs, especially in the Cards condition, and the PCQs, especially in the Board condition.

Overall, based on the scores of the participants in the present study, AS appears to be the score with the highest feasibility, out of the three main scores of the Alien Game. This is based on the finding that it had the widest range of scores and it did not have any floor or ceiling effects. It is a measure based on a child's elimination strategy, therefore it has good potential to indicate their concept formation abilities.

Table 5

Descriptive statistics and Wilcoxon Singed Rank Test statistics for the Alien Game scores in the Cards versus Board conditions.

						Wilcoxon		
		Mean	SD	Min.	Max.	W	Z	Exact Sig.
IAS	Cards	6.23	4.73	0	12	3.33	-0.68	.53
	Boards	7.23	4.11	0	12			
AS	Cards	11.54	6.85	3	20	5.43	-1.07	.31
	Boards	13.31	5.72	3	20			
WAS	Cards	2.92	1.80	1	5	5.00	-1.31	.21
	Boards	1.77	1.54	1	5			
TNQs	Cards	7.31	2.50	4	11	6.00	-1.88	.06
	Boards	9.54	2.33	5	11			
CQs	Cards	2.77	2.17	0	5	3.00	-2.01	.06
	Boards	3.46	1.98	0	6			
PCQs	Cards	3.00	4.24	0	10	2.00	-0.96	.50
	Boards	2.08	3.59	0	10			
RedQs	Cards	1.08	1.61	0	5	4.00	-2.21	**.03**
	Boards	2.69	2.21	0	6			
RepQs	Cards	0.15	0.55	0	2	3.20	-1.19	.38
	Boards	0.62	1.12	0	4			

Note. IAS= Initial Abstraction Score, AS=Abstraction Score, WAS= Weighted Achievement Score. TNQs=Total Number of Questions, CQs= Number of Conceptual Questions, PCQs= Number of Pseudo-Conceptual Questions, RedQs= Number of Redundant Questions, RepQs= Number of Repetitive Questions. N=13. *P* values represent 2-tailed significance

3.4 Consideration of Acceptability of the Alien Game

All participants (N=14) were asked to rate the Alien Game along a child-friendly 5-point Likert Scale (Please see Appendix 6.2). The last seven participants were asked to rate Similarities and Matrix Reasoning. As shown on Table 6, the majority of participants rated the Alien Game more favourably than Similarities and Matrix Reasoning. This indicates that the format of the Alien Game had high levels of acceptability by the participants, meaning that they found it engaging and enjoyable.

Table 6

Likability of the Alien Game, Similarities and Matrix Reasoning

	Alien Game (N=14)	Similarities (N=7)	Matrix Reasoning (N=7)
Fantastic	86%	43%	43%
Really Good	7%	43%	14%
Okay	7%	0%	43%
Not very good	0%	0%	0%
Awful	0%	14%	0%

Note. Ratings presented in terms of percentage of the N.

3.5 Comparisons between descriptive statistics of Cards and Board conditions

One of the research questions of the present study was whether there would be a difference in participants' performance in the Cards versus the Board condition. In the Cards condition, eliminated aliens are flipped over, and this may help to reduce the Working Memory load of the task as participants do not need to keep track of which of the aliens have been eliminated or to remember which questions they have already asked.

A visual inspection of the descriptive statistics of the scores in the Cards and Board conditions (please see Table 5) was conducted to review the trends in the data. The IAS and AS scores were slightly higher in the Board condition, therefore it appears that participants were asking questions with slightly higher MEPs in the Board condition. However, judging by the higher WAS in the Cards condition, it appears that the number of questions participants asked in the Cards condition was closer to the optimum number of questions (as explained in Table 2). The Standard Deviations of the scores were large, therefore there was a significant overlap of scores in the Board and Cards condition. A Wilcoxon Signed Ranks Test suggested that the differences in the IAS, AS and WAS scores was not statistically significant. The Exact Tests resampling procedure available on SPSS was used to yield more robust results.

With respect to the types of questions asked, as shown on Table 5, it appears that participants asked on average more questions in total in the Board Condition. More specifically, they asked more conceptual questions,

as well as more redundant and repetitive questions. The only opposite trend was observed for the Pseudo-conceptual questions, which were on average lower in the Boards conditions. Similar to the IAS, AS and WAS, the Standard Deviation of the averages in the type of question scores were large and reflected the overlap between values. However, a Wilcoxon Signed Ranks Test found that the difference in redundant questions was statistically significant. The p value was marginally above significance for the CQs and TNQs scores, which was further corroborated by the Z-score for the CQs score.

The comparison of the Cards versus Board scores with a small sample can only allow for tentative understanding of the data. However, it appears that in the Board condition, participants had the tendency to ask more questions overall, and more conceptual questions specifically. They tended to achieve higher MEPs in their initial and first three questions.

3.6 Consideration of Concurrent Validity of the Alien Game scores

One of the research questions of the present study was to explore the extent the Alien Game scores would correlate with established concept formation measures, specifically the WISC-IV Similarities and Matrix Reasoning. Spearman's Rank Correlations were used to explore this. The scores from the Cards and Board conditions were explored separately.

For this analysis only a subset of the available Alien Game scores was used. The first was the AS, given the width of scores obtained and its theoretical significance of capturing the MEP of the participant's first three questions, i.e.

the quality of the participants' strategy. The second score that was included in the subsequent analysis was the CQs, because of its theoretical significance as a measure of participants' ability to utilise the alien attributes to ask elimination questions.

The IAS was not used due to its high correlation with the AS (please see Table 11 and Table 12 in Appendix 6.6) This was not surprising as the AS consists of the IAS (i.e. the MEP of the first question) plus the MEP from the second and third question. The WAS was not selected due to its narrow range of scores. The TNQs, PCQs, RedQs and RepQs were not selected due to their narrow range of scores and the fact that theoretically they were not relevant for the exploration of the Alien Game's concurrent validity.

As shown on Table 7, all scores had a strong positive correlation with Similarities. The Cards CQs had the strongest, followed by Board AS, Board CQs, and Cards AS. Three scores had a moderate positive relationship with Matrix Reasoning. Board AS had the strongest, followed by Cards AS and Board CQs. Overall, it appears that AS and CQs have good concurrent validity with Similarities, and some moderate concurrent validity with Matrix Reasoning.

The AS and CQs scores did not correlate with Age, apart from Board AS that had a moderate negative relationship with Age. This means that the younger participants were more likely to obtain a higher AS score in the Board Condition.

Table 7

Spearman's Rank Correlation coefficients for the relationships between Cards and Board AS and CQs scores with Age, WISC-IV Similarities and Matrix Reasoning Raw Scores and CHEXI Working Memory, and Inhibition Raw Scores.

		WISC-IV		CHEXI	
	Age (in months)	Similarities	Matrix Reasoning	Working Memory	Inhibition
Cards AS	-.202 (.508)	**.515** (.072)	**.364** (.222)	**-.306** (.309)	-.144 (.639)
Board AS	**-.373** (.209)	**.564** (.045)	**.412** (.162)	-.186 (.543)	-.138 (.653)
Cards CQs	-.188 (.540)	**.618** (.024)	.296 (.326)	**-.311** (.300)	**-.308** (.306)
Board CQs	-.188 (.539)	**.523** (.067)	**.309** (.304)	-.288 (.339)	-.266 (.380)

Note. Moderate and large effect sizes are in **bold**. AS=Abstraction Score. N=13. *P* values are 2-tailed and are presented in brackets.

3.7 Consideration of Predictive Validity of the Alien Game scores

Further analysis was conducted to evaluate the predictive validity of the AS and CQs scores in terms of their relationship to the CHEXI Working Memory and Inhibition scores. Given its theoretical significance and that it was the only variable that was significantly different in terms of Cards and Board conditions, the RedQs were included in this analysis. The RepQs were not included as they were very limited. The reason that only Working Memory and Inhibition were included from the CHEXI, is due to their theoretical significance in concept formation processes. In addition, CHEXI Working Memory and Inhibition are the scores that are best supported by factor-analytic studies (Thorell & Nyberg, 2008).

As shown in Table 7, Cards AS had a moderate negative correlation with Working Memory, and did not correlate with Inhibition. Board AS did not correlate with either Working Memory or inhibition. Both Cards and Board CQs had a moderate negative correlation with both Working Memory and Inhibition. Neither Cards and Board RedQs scores had a correlation with Working Memory or Inhibition.

Overall, the tentative interpretation is that the Cards CQs showed moderate predictive validity of Working Memory and Inhibition, and that Cards AS score showed moderate predictive validity of Working Memory.

3.8 Observations of trends in Children's responses

A review of the children's questions was conducted to gain additional insight into the data. The aim was to better understand which alien attributes were asked about the most, and what the nature of the redundant questions was.

As shown in Table 8, In the Card condition, the most popular attributes in order of frequency were: colour, wings, legs, eyes, eyebrows, antennae and tails. Nose, teeth, shape, and horns were the attributes least asked about. Arms did not feature in any of the questions. In the Board condition, the most popular attributes in order of frequency were: colour, legs, wings, eyes, tails, antennae, ears and horns. Shape, arms, teeth and eye brows were the attributes least asked about. Nose did not feature in the any of the questions.

As shown in Table 9, the vast majority of redundant questions were due to children asking questions based on attributes that were shared by all remaining aliens. For example ,asking if an alien had wings, when all remaining aliens had wings. This trend was observed in both Cards and Board Conditions. Other types of redundant questions that were observed were confirmatory questions of trichotomous attributes, e.g. participants asking if the alien was blue, after they had found out it was neither black nor yellow. In both conditions, there was a spatial-based question (e.g. "Is the alien on the left side?") and a question whether the alien was among the first ten aliens. These were asked by the same two children. Finally, a single participant asked a number of questions that were based on the aliens' hobbies (e.g. "Do they like to read books?").

In the Board condition, another reason for the higher number of redundant questions was that children failed to keep track of which aliens they had already eliminated. As a result they continued to ask questions beyond the point that they had enough information to have found the target alien. This affected four of the participants. Overall, in the Cards condition 10 participants found the target alien versus 5 participants in the Board condition.

3.9 Analysis of Cultural Loadings of the Alien Game

The number of languages spoken at home was the only variable that could be used to support a statistical analysis (e.g. General Linear Model) of the effect of culture on the Alien Game scores. However, due to the small N, the three possible Language groups would be too small (One language: N=4; Two languages: N=5; Three languages: N=4). For this reason, this aspect of the Alien Game scores were not explored in the present study.

Table 8

Frequency of questions asking about a specific attribute.

Attribute	Frequency of occurrence		
	Cards	Board	Total
Colour	9	14	23
Number of legs	6	11	17
Wings	8	7	15
Number of Eyes	6	6	12
Number of Antennae	4	5	9
Tail	3	5	8
Ears	2	4	6
Horns	1	4	5
Eyebrows	4	1	5
Shape	2	3	5
Teeth	2	1	3
Spatial	1	2	3
Arms	0	2	2
Nose	2	0	2
Alien Order	1	1	2

Table 9

Redundant Questions' Reason for Failure to Eliminate

	Frequency of occurrence		
	Cards	**Board**	**Total**
Attribute is shared by all remaining Aliens	13	20	33
Redundant confirmatory question	0	5	5
Question not based on visible attribute	4	0	4
Alien already eliminated	0	4	4
Not appropriate Yes/No format	1	0	1

4. DISCUSSION

4.1 Summary of Research Questions and Main findings

4.1.1 Revisiting the Aims, Research Questions and the Epistemological Position

The aim of the present study was to explore the downward extension of the Alien Game as a child-friendly concept formation measure for six and seven year olds. The research questions were:

- whether the Alien Game has good applicability and acceptability when used with this age group
- whether the Alien Game scores showed good concurrent validity with established concept formation measures
- whether the Alien Game scores showed predictive validity for Working Memory and Inhibition scores
- whether performance on the Alien Game is affected by cultural factors
- whether there would be any difference in performance between the Board and Card conditions of the Alien game, which according to the theory of concept formation were expected to differ in terms of their working memory and inhibition demands.

Each of these questions will be considered in relation to the results of the present study. They will be discussed from the epistemological position of critical realism. Critical realism views the terminology used in the literature,

and in the present study, as indirect representations of entities that are beyond our direct perception. From that perspective, the Alien Game scores are only an approximation of a specific cognitive ability model that we have created to enable us to talk about cognition. What is important is that we pay attention to all the factors that may influence performance in concept formation tests, as this understanding will help us improve our models of concept formation and ultimately of cognition.

In addition, it is important to maintain awareness of the fact that caution is needed throughout the interpretation of the results of the present study given the limitation of the small sample size, and where mean values are compared, the overlapping range of values.

4.1.2 Applicability and Acceptability of the Alien Game

All the Alien Game scores were considered for their applicability to measure concept formation. This included the Abstraction Score, the Initial Abstraction Score, the Weighted Achievement Score, and the number of questions asked, i.e. Total Number of Questions (TNQs), number of Conceptual Questions (CQs), Pseudo-Conceptual Questions (PCQs), Redundant Questions (RedQs) and Repetitive Questions (RepQs).

Of all the Alien Game variables, the AS had the wider range of values obtained, and had no floor or ceiling effects. The range of possible values of the AS is based on the range of the Minimum Elimination Potential (MEP) that can be achieved in the first three questions of a trial. The lowest end is 0 if no CQs or PCQs are asked. The highest MEP that can be achieved is if the child asked three CQs in a row. The revised design of the aliens in the

present study ensured that there was an equal allocation of attributes spread among them. There were seven attributes with two variations (dichotomous attributes) and six attributes with three variations (trichotomous attributes). The best AS can be achieved if participants ask as many CQs based on dichotomous attributes as possible (obtaining an MEP of 50% or as close to 50% as possible) meaning that regardless of the participant's question being answered with a yes or a no, at least half the remaining aliens will be eliminated with each question. Given the number of aliens (24) and the number of dichotomous questions, this is possible for the first three questions, which are covered by the AS. It is therefore possible for AS to be as high as 21. As the number of remaining aliens reduces, the allocation of remaining attributes become more irregular, meaning that by the fourth question it is less possible for participants to ask CQs with a MEP as high as 50%. However, this does not affect the AS score.

The IAS and WAS had a more limited range than the AS. For the IAS, the possible MEP scores can be 0 if the child asks a question that is redundant, 1 if the child asks a PCQ, 8 if thy ask a CQ based on a trichotomous attribute or 12 if they ask a CQ based on a dichotomous variable. For the WAS the score can be between 1 to 5, depending on how close the total number of questions is to the number of questions that are needed to find the target alien if the child follows an optimum strategy. The optimum number of questions was 5, as adapted from Pavitt (2017).

From the Alien Game scores based on types of questions, the CQs were most theoretically relevant as concept formation measure. A high CQs number indicates that the child had the ability to formulate questions based on conceptual categories of the aliens. However, the highest number of CQs

does not necessarily reflect the highest concept formation ability, as it may reflect a poorer conceptual strategy, especially if it goes beyond the five questions considered to reflect the optimum strategy. As a result, CQs may not be as applicable as measuring concept formation as the AS is. For the AS, the higher the score the better, which makes it a clearer reflector of concept formation.

The TNQs and PCQs had a relatively wide range. However, neither are considered applicable as concept formation scores, due to the fact that that a high or low score does not necessarily reflect a strategy based on concept formation. A low score of PCQs could be the result of a high number of RedQs and RepQs. The RedQs and RepQs have a similar theoretical limitation, and in addition their ranges were narrow.

Overall, the AS was the score with the qualities needed to give it high applicability as a concept formation score for six and seven year old children. From a practicability point of view, the AS is based on only the three first questions of a child, making it easy to calculate (an important point when considering ease of use of the Alien Game for busy clinicians).

In terms of the Alien Game's acceptability and child-friendliness, it appeared to be favoured by the vast majority of participants, with 86% of participants giving it a 5/5 rating, and none rating it below 2/5. This is positive evidence for the Alien Game in terms of its aim to be child-friendly and enjoyable. The feedback survey was limited and conducted in an already small sample, and did not include qualitative questions asking participants to explain their scores. Therefore there is no data to suggest which aspects of the Alien Game participants liked or if there was anything that could be improved. For these reasons, these findings should be interpreted with caution. However,

the high level of enjoyment reported indicates a high level of overall engagement, which means that the scores obtained by the present study are likely reflecting the children's genuine performance on the Alien Game.

Although the aim of the likability survey was to assess acceptability of the Alien Game, it was decided that it would be meaningful to have participants' ratings for the WISC-IV Similarities and Matrix Reasoning to make comparisons. Although this data was available for a subset of participants, it provided some evidence that the Alien Game was more enjoyable for the children than either Similarities or Matrix Reasoning.

4.1.3 Concurrent Validity of the Alien Game

The WISC-IV Similarities and Matrix Reasoning tasks are established measures of verbal and non-verbal concept formation respectively (McGee et al., 2008). They were used to evaluate their correlations with the AS and CQs scores, which would be an indication that these Alien Game scores did measure concept formation. Interpretations of these relationships need to be made with caution, in the context of the small sample size of the present study.

Both AS and CQs scores were found to have a strong positive relationship with Similarities. Therefore this is evidence of good concurrent validity for measuring verbal concept formation. The relationship with Matrix Reasoning of these variables was moderate, and this is evidence of a moderate concurrent validity for measuring non-verbal concept formation. The Alien Game is intended as a predominantly visual task, therefore it is a surprising finding that concurrent validity with Similarities was the higher of the two.

Perhaps this can be understood by looking at the design differences between the Alien Game and Matrix Reasoning. In Matrix Reasoning, participants are shown a visual pattern, along with a number of choices, one of which fits the pattern. This means that the correct option is visible to them and they can solve the task through perceptual reasoning and mental rotation. By contrast, in the Alien Game participants are asked to group and categorise a large number of objects, and this is where they may use semantic categories more than perceptual ones.

An important limitation of the data is that the sample was found to have an overall score of Matrix Reasoning that was below the general population and this may impact the analysis of the relationship between the Alien Game scores to Matrix Reasoning.

The CQs scores were the least consistent in terms of their concurrent validity. Cards CQs had the highest correlation with Similarities and the lowest correlation with Matrix Reasoning. The Board CQs had the second lowest relationship with Similarities and the lowest correlation with Matrix Reasoning. This is an indication that asking a higher number of CQs may be more associated with verbal concept formation, and potentially with general verbal ability, than it is a reflection of non-verbal concept formation.

The finding that the AS scores had the highest correlations with Similarities and Matrix Reasoning adds more weight to the argument that AS may be the most suitable Alien Game score to measure concept formation, although it may be more representative of verbal than non-verbal concept formation.

4.1.4 Predictive Validity of the Alien Game

The AS and CQ scores of the Alien Game were found to have a strong relationship with the CHEXI working memory and inhibition, meaning that they have low predictive validity for working memory and inhibition. Only the Cards AS and Cards CQs relationships with working memory had a coefficient that was marginally above the criterion for a moderate effect size. Only the Card CQs relationship with inhibition had a coefficient that was marginally above the criterion for a moderate size effect.

4.1.5 Cultural Fairness of the Alien Game

It was not possible to explore this question, given the limitations of the data. The sample size was too small to enable meaningful statistical comparisons of the effects of cultural factors on the Alien Game performance. In addition, the demographic information on languages spoken by the participants was not clear in terms of how much any additional languages were used at home, or other objective measurement such as range of vocabulary of additional languages.

More specifically, the sample of the 13 participants could not be meaningfully divided into even cultural categories on the basis of primary languages spoken at home or the number of languages spoken at home. If the number of languages spoken at home would have been used as a culture variable, it would have produced a group of four participants speaking one language, a second group of five participants speaking two languages and a third group of four participants speaking three languages. However, it would have been very difficult to know what the variable means in terms of language knowledge and use as there were no ways to measure that and it could be

very different from one participant to the next. The second variable that could have been used in a slightly more objective manner, is the primary language of the participants. However, this would have resulted into three uneven groups, as 10 participants spoke English as their primary language, one participant stated their primary language was not English and two participants stated they could not distinguish between English and another language as their primary language. As a result, this variable could not be used for any meaningful comparisons between groups.

Given the significance of the cultural fairness aspect of the Alien Game, it was deemed inappropriate to attempt any inferential or descriptive analysis to avoid the risk of misinterpreting or overinterpreting the data.

4.1.6 Comparison between Board and Cards Conditions

In the present study two different formats of the Alien Game were used. In the Cards condition, eliminated aliens are no longer visible, reducing both the need to remember which questions have been asked, and to keep track of which aliens have been eliminated and which aliens remain. In the Board condition, all aliens are visible all of the time, meaning that tracking eliminated and non-eliminated aliens has to be done using working memory. Therefore the main difference between the two conditions was the need for working memory input.

There were two main aspects of the data that were relevant to the comparison between the Cards and Board conditions. The first aspect was the performance of participants in relation to the Alien Game scores. The

second aspect was the informal qualitative observations of the researcher in terms of the questions that children asked.

With respect to Alien Game scores, the only statistically significant result was that, in the Board condition, participants asked more redundant questions. This is not surprising, as the Cards condition was intended to reduce these types of questions.

The review of the questions participants asked indicated an interesting trend, related to the above result. In the Board condition, four participants asked more questions than were needed to find the target alien. This was due to their difficulty keeping track of which aliens had already been eliminated, and not noticing when they had enough information to find the target alien. Any questions beyond that point were marked as redundant, and this may have resulted in the significant difference between the Board RedQs and Cards RedQs. However, it is surprising that participants did not ask more RedQs and RepQs, and that despite these they obtained higher average IAS and AS scores.

The most prevalent type of RedQs, in both the Cards and the Board conditions, was questions based on attributes that were shared by all remaining aliens. This suggests that participants struggled with identifying the most pertinent attributes to based their questions on, although it is not clear why. The fact that this happened more often in the Board condition may be an indication that this type of error was related to the difficulty in keeping track of which aliens had been eliminated. However, the fact that these errors occurred frequently in the Cards condition indicates that there was a second reason, unrelated to working memory. One possibility could be that, although participants had the procedural knowledge, i.e. how to ask the yes/no

questions to play the Alien Game, they did not have a developed enough conceptual knowledge, i.e. why they were asking the questions they chose (Schneider et al., 2011). Without the support of conceptual knowledge, participants would be less likely to have developed the procedural flexibility to turn their attention to the pertinent attributes to eliminate more non-target aliens in a strategic way (Schneider et al., 2011).

Despite the challenges of the Board condition, there was no statistical difference between the Cards and Board AS. Without replicating this finding with a larger sample, it is hard to know if this finding reflects something about the role of working memory in completing the Alien Game.

Any conclusions from the comparison of the Cards and Board conditions need to be made with a degree of caution, given the small size of the sample, the wide error margins as shown by the Standard Deviation values and the fact that the comparison is based on a single trial of each condition per participant.

In the literature, the role of working memory in concept formation is not clear and there seem to be two competing theories. One suggests that a high working memory load has a negative impact on concept formation (Crone & Van der Molen, 2004), and the other suggests that it activates concept formation, out of the need to organise the various objects that must be held in its limited capacity (McGonicle-Chalmers & Alderson-Day, 2010). However, they may refer to different types of working memory.

Despite the sample limitations, theoretically, it may be possible that in the Cards condition, participants benefitted from the removal of the eliminated aliens, but could not benefit from the Principled sorting Heuristic. Conversely,

in the Board condition, participants benefited from the Principled sorting heuristic but were disadvantaged by the working memory load associated with this condition.

From a critical realist perspective, it is important to be aware that participants may have been acting using mental processes not defined by the theorised constructs of concept formation, working memory or inhibition. From this epistemological position, it would be enlightening to have more information from the participants themselves as to how they were solving the task. This, and similar qualitative approaches, could increase the understanding of the relationship between their scores and the types questions which were seen as errors in the Alien Game.

4.2 Comparison with Pavitt (2017)

The results of the present study cannot be directly compared with those of Pavitt (2017), due to the difference in how the scores were calculated. In Pavitt (2017) participants completed four trials of the same format of the Alien Game, whereas the data available from the present study are from a single trial of the Cards condition and a single trial of the Board condition. Despite these differences it is interesting to consider any similarities and points of divergence from the trends and patterns in the Pavitt (2017) study.

One difference between the present study and Pavitt (2017) is in the way the Alien Game's scores correlated with Similarities and Matrix Reasoning. In Pavitt (2017), the strongest correlations were between Similarities and the WAS and IAS scores, and both coefficients were marginally below the

criterion for a moderate relationship. In addition, WAS had a negative moderate correlation with Matrix Reasoning Raw scores. By contrast, in the present study the AS was selected to check concurrent validity. The AS achieved a stronger relationship with Similarities and Matrix Reasoning than what was found by Pavitt (2017). In addition, the WAS did not correlate with Similarities or Matrix Reasoning (please see Table 13 in Appendix 6.6). It is not clear what gave rise to these differences. One possibility is that the equal distribution of alien variables in the present study may have changed the way children responded to the task, therefore generating different patterns in the scores. In addition, as the age groups in the two studies are different, it may reflect differences in the type and quality of the strategies of six and seven year olds.

A similarity with Pavitt (2017) is that in both studies the Alien Game scores had stronger correlations with Similarities than with Matrix Reasoning. This is further evidence that verbal concept formation may play a more important role than non-verbal concept formation when playing the Alien Game.

4.3 Critical Evaluation

4.3.1 Strengths of the present study.

The main strength of the present study was its ability to identify that the AS score has high concurrent validity with Similarities and Matrix Reasoning, both established concept formation measures, suggesting that the AS is a measure of concept formation. The fact that its game-based design was very highly rated is an indication that children were likely to be highly engaged in

the alien-finding task, and that their performance in the game was a true reflection of their ability. This indicates that the Alien Game may have high ecological validity. The adaptation of the Alien Game used in the present study contained an experimental pilot condition that sought to reduce the impact of working memory demands on children (the Cards condition). In addition, the administration of the Alien Game follows the advice in the literature re reducing the need for children to manipulate stimuli, as it is neither pen/paper- or computer-based. All of these factors are strengths in helping the Alien Game eliminate confounding variables that could interfere with measuring concept formation.

The present study also made improvements to the stimuli. New aliens were designed, with their features equally distributed to eliminate the element of luck in how many non-target aliens could be eliminated by each question. In addition, the features of the aliens were made clearer and more visible. The colours of the aliens were kept the same as in the Pavitt (2017) study, in order to make the Alien Game accessible to any children that experience colour-blindness.

As a sample, participants appeared to be representative of the general population as far as their Similarities Scaled scores were concerned, as their average (mean) performance was 10, which is the population's expected average.

Through the epistemological position of critical realism, the present study sought to understand the role of cultural factors in participants' performance in the Alien Game. It is important to contribute to cognitive assessment research from a critical realism point of view, to create a counter-narrative to the realist and positivist stances on cognition that can enable endemic

oppression of disadvantaged groups. It is therefore a strength of the present study that it holds awareness that the scores analysed and discussed are not a direct representation of an objectively true cognitive ability. Instead they represent a model of a cognitive ability that has been constructed as concept formation, which may ultimately and subjectively be very different to our external perception of it. This awareness allows for more openness to understanding how the Alien Game performance may be the result of a number of different factors, including socio-cultural, and not only cognitive ones.

4.3.2 Limitations of the present study

The limitations of the present study revolve around methodology, data analysis and the sample size.

Two main limitations of the study were the small sample size (fourteen participants), and that there were missing data. The small sample size means that caution is needed when drawing conclusions from the findings, and that no analysis could be performed to speak to the cultural loadings of the Alien Game, meaning that one of the original research questions cannot be answered.

In relation to missing data, children had difficulties answering some of the demographic questions, such as the professions of their parents. It is possible that the study did not take into account the level of information that children of that age could provide, and it is a limitation that the study did not create additional channels for obtaining demographic information from parents.

Other missing data were caused when some children asked to start the Similarities and Matrix Reasoning tasks before they had completed all of their Alien Game trials. The missing data meant that the Alien Game scores were based on a single trial of the Cards and Board conditions, when Pavitt (2017) calculated her scores based on four trials. This may have affected the distribution of the data, as well as the way they correlated with the other variables. The results, therefore, need to be interpreted with caution.

The design of the Alien Game did not incorporate ways to explore qualitative elements, such as participants' understanding of the instructions, their reasoning while engaging with the task, detailed feedback on the aspects of the task they found enjoyable, and what could be done to improve enjoyability of the Alien Game.

As there were no trial rounds of the Alien Game procedure, children did not have a chance to practice and to demonstrate if they understood the Alien Game before the actual task began. The small sample did not allow for a deeper understanding of the performance of children who relied on different strategies. For example, it was not possible to meaningfully compare children who made no errors to those who made errors. A larger sample, in combination with a practice trial and more qualitative information on participants strategy could have helped to group participants in ways to facilitate such comparisons.

It is not known if the children gave the Alien Game positive ratings because they knew from the consenting information that the Alien Game was important to the researcher, and therefore they showed desirability bias towards it.

Finally, the sample was not random. It was a convenience sample, and therefore may be limited in terms of how well it represented the average population.

4.4 Directions for Future Research

The present study adds more evidence to the potential of the Alien Game to be used as a game-based measure of concept formation. The results from the present study showed that the Abstraction Scores (AS) of six and seven year old children had a strong correlation with an established verbal concept formation measure and a moderate correlation with a non-verbal concept formation measure. Although this finding is based on a small sample size and should be interpreted with caution, it means that further research into the Alien Game is justified, and the following steps and directions are recommended by the researcher.

It is important to replicate the results of the present study with a larger sample. It would be helpful to explore the test-retest reliability by testing the same group of children at two different times.

Future research could explore if the high number of alien attributes is necessary for six and seven year olds, as this could be another way to make age appropriate adaptations to the Alien Game. An informal review of the frequency at which specific attributes were referenced by participants in their questions showed that the arms, nose and teeth were used the least. This may mean that they are not as perceptible and therefore could be revised or removed.

A further suggestion for material development could be that, if a Cards condition is kept, to cross out the eliminated aliens instead of flipping over the cards. This would save time from an administration point of view (as it is time consuming to flip and then reset the cards to use them in another trial). However, if the “cross-out” condition of the Alien Game is used, it would be interesting to compare it to the Cards and Board conditions. It is conceivable that in this third condition, working memory load is kept low but that the Principled Sorting Heuristic remains activated, as all aliens are visible, even when they are crossed out. It would be interesting to see if this would have an impact on children’s performance in the Alien Game. If not, then this would be more evidence that working memory does not influence performance in the Alien Game.

Qualitative research may help to make further improvements in increasing the child-friendliness and developmental suitability of the Alien Game. Some of the areas to be explored qualitatively would be the types of questions children asked, as well as their reasoning for asking those questions. It would be helpful to explore other aspects of the Alien Game, such as how clear the instructions were and, more generally, if participants felt that there are elements that could be improved.

The Alien Game is intended to be a culturally fairer measure of concept formation. Therefore future research is needed in order to explore cultural loadings of the Alien Game, especially as the aim for the Alien Game is to be a standardised measure. Participants’ language and culture could be used in regression analyses to evaluate equivalence, i.e. that the different cultures yield similar loadings (Meredith, 1993). In a similar way, socio-economic

factors, such as parental occupation, could be examined in relation to possible effects of performance in the Alien Game.

If the results of the present study are replicated sufficiently, and the cultural loadings of the Alien game are appropriately explored, then the next steps of development could be to explore the utility of the Alien Game with six and seven year old children who have specific neurodevelopmental diagnoses, e.g. autism or brain injury. Ultimately, the Alien Game, as a concept formation measure, would be helpful for assessing (and thus securing targeted support for) clinical populations. However, it will be very important for further research to first assess validity and acceptability of the Alien Game with children with these diagnoses (Allen et al., 2006).

4.5 Reflections

In the context of the critical realist epistemological position, it is important that space is given to reflect on the interaction of the agency and identity of the researcher with the structures and context around the present study (Flanagan, 1981). I will make Ethical and Professional Considerations, as well as personal reflections.

First of all, it is important to reflect on the weight of responsibility when designing or contributing to the development of a tool that purports to measure cognitive ability. Through a critical realism stance, I have tried to highlight that although such tools have merit and can be helpful, at the same

time, they hold a lot of power, and therefore should be designed and tested carefully.

When I considered the Alien Game as my doctoral thesis project, I felt enthusiastic about the wider goal of designing a tool that is both child-friendly, but also culturally fairer. However, in a dilemma that I continue to reflect on, my mere engagement with this type of research means that I give strength to positivist and realist positions by designing a tool that aims to result in a standardised measure for typically developing children and children with developmental diagnoses and brain injury. It is difficult to disentangle a critical realism-informed understanding of the data and results from contexts heavily influenced by Western medical, clinical and positivist narratives. For example, although I believe in supporting children who need educational support, I do not believe that a low score in the Alien Game should be translated as that child having a concept formation 'difficulty', or 'deficit'. I envision the Alien Game to be a measure of performance within that specific situation, and not of inherent ability. In fact, I have reflected that maybe a goal of further research would be how to incorporate feedback into the Alien Game design to explore the features of feedback that may enable children to learn through playing the Alien Game and whether they can develop increasingly better conceptual strategies.

Despite these thoughts, there is a large part of me that considers it important to design a measure that would have utility in a clinical context. However, I hope that the Alien Game will continue to be situated in a critical interpretation of children's performance.

Through reading cultural fairness literature, I have become increasingly aware of my identity as a white, non-British researcher. This means that I am

familiar with the idea of cognitive measures and have my own test-wiseness, but narrows my own understanding of how the Alien Game may be perceived by children from cultures where the concept of cognitive testing is not familiar or considered as relevant. For example, the instruction that children had to find the target alien as quicky as possible may make sense to me in my own culture, but it may not do so in all cultures.

As I read about the interaction of the cultural identity of the person administering the test and person being tested (Van de Vijver & Poortinga, 1992), I reflected on how methodologically complex it is to measure the effects of culture on performance in cognitive tests, due to its many subtle, multi-faceted and nuanced manifestations. This is another reason for maintaining a critical realist stance when interpreting my data, and highlights the need for further qualitative, as well as quantitative, research into understanding such complex cultural influences.

It is important, from an ethical as well as a professional point of view, that people should be offered measures that are culturally well-informed. If we abandon this commitment, and take an unquestioning positivist stance on the interpretation of performance in assessment tools such as the Alien Game, problematic and unethical applications may continue to occur, replicating patterns of injustice and discrimination against already disadvantaged groups.

4.6 Theoretical and Practical Implications

The implication of the present study is that the Alien Game can be used as a measure of concept formation for six and seven year old children. The Alien Game AS score may be a good measure of concept formation in six and seven year old children. This score is based on the sum of the Minimum Elimination Potentials of the first three questions participants ask. It therefore measures their ability to choose which attributes they should base their questions on in order to achieve an optimum elimination strategy. Another implication is that a single trial of the Alien Game may be sufficient to obtain a child's concept formation score.

The Alien Game was rated highly by children in the age group and therefore may be a better way to engage children than other tools (such as Similarities and Matrix Reasoning), thus enabling a better and more valid measurement of concept formation. In addition, the Alien Game lends itself for further development and research in terms of its cultural fairness. From a critical realism point of view, this means that the Alien Game could help to improve existing theoretical understanding of concept formation, by providing a more accurate model of children's concept formation.

The mitigation of confounding variables through the Alien Game's design means it may have utility for measuring concept formation for children in a clinical context, for example in children with neurodevelopmental diagnoses, or brain injury. The availability of child-friendly measures may enable a more accurate assessment of concept formation in these children, and therefore better provision of targeted support.

4.7 Overall Summary and Conclusion

The Alien Game is a child-friendly game-based task initially developed by Pavitt (2017), where children eliminate aliens by asking yes or no questions, based on the aliens' visible features, in order to find a target alien. The resulting score is seen as measure of concept formation, as this cognitive function would be required to develop the best strategy to find the target alien. In addition to being child-friendly, the Alien Game is intended as a culturally fairer measure of concept formation.

The focus of the present study was to assess whether the Alien Game can be used as a concept formation measure for six and seven year-old children. On the basis of a small sample of participants, the results of the study indicated that the Alien Game can be used with this age group, and it was rated highly in terms of its acceptability by the participants. The Abstraction Score of the Alien Game was found to have moderate to good concurrent validity with established concept formation measures. There was no difference between a condition that was designed to reduce the working memory load and a condition that was expected to have a higher working memory load. It was one of the aims of the study to explore effects of cultural factors, such as the primary language of participants, on performance but this was not explored due to the small sample size.

Overall, the findings of the present study suggest that the Alien Game merits further development as a concept formation measure for six and seven year old children. Further development should involve replication with a larger

sample, and analysis of the relationship between culture and performance in the Alien Game, in accordance with its aim to be a culturally fairer test of concept formation.

5. REFERENCES

Adobe Inc. (2022). Adobe Illustrator, Version 23. https://www.adobe.com/products/illustrator.html

Albert, D., Opwis, K., & Regard, M. (2010). Effect of drawing hand and age on figural fluency: A graphomotor study with the Five-Point Test in children. *Child Neuropsychology, 16*(1), 32–41. https://doi.org/10.1080/09297040903049061

Allen, D. N., Knatz, D. T., & Mayfield, J. (2006). Validity of the Children's Category Test-Level 1 in a clinical sample with heterogeneous forms of brain dysfunction. *Archives of Clinical Neuropsychology, 21*(7), 711–720. https://doi.org/10.1016/j.acn.2006.08.003

Anastasi, A. (1968). *Psychological testing.* MacMillan Co.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45,* 153–219. https://doi.org/10.1016/S0010-0285(02)00002-6

Arvey, R. D. (1972). Some comments on Culture Fair Tests. *Personnel Psychology. 25*, 433–446.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.

Baldo, J. V., Delis, D. C., Wilkins, D. P., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Archives of Clinical*

Neuropsychology, 19, 407–419. https://doi.org/10.1016/S0887-6177(03)00074-X

Bechara, A., Damasio, A. R., Damasio, H & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7–15.

Bechara, A., Damasio, H., Tanel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295

Bello, D. T., Allen, D. N., & Mayfield, J. (2008). Sensitivity of the children's category test level 2 to brain dysfunction. *Archives of Clinical Neuropsychology, 23*(3), 329–339. https://doi.org/10.1016/j.acn.2007.12.002

Berry, J. W. (1994). Acculturation and psychological adaptation: An overview. In A. Bouvy, F. J. R. Van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 129–141). Swets & Zeitlinger.

Bhaskar, R. (1998) General Introduction. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Reading*. Routledge.

Birney, D. P., Halford, G. S., & Andrews, G. (2006). Measuring the Influence of Complexity on Relational Reasoning: The Development of the Latin Square Task. *Educational and Psychological Measurement, 66*(1), 146–171. https://doi.org/10.1177/0013164405278570

Blaye, A., & Jacques, S. (2009). Categorical flexibility in preschoolers: Contributions of conceptual knowledge and executive control. *Developmental Science, 12*(6), 863–873 https://doi.org/10.1111/j.1467-7687.2009.00832.x

Booth, A.E., Schuler, K., & Zajicek, R. (2010). Specifying the role of function in infant categorization. *Infant Behavior Development. 33*, 672–684. https://doi.org/10.1016/j.infbeh.2010.09.003

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology, 46*(2), 350–365. https://doi.org/10.1037/a0018411

Brookman-Byrne, A., Mareschal, D., Tolmie, A.K., & Dumontheil, I. (2019). Verbal analogical Reasoning and Nonverbal Matrix Reasoning to Science and Maths Problem-Solving in Adolescence. *Mind, Brain and Education, 13*, 211–223.

Bruner J, Goodnow J, & Austin G. (1967). *A study of thinking*. Science Editions.

Bryman, A. (2008). *Social Research Methods* (3rd ed.). Oxford University Press.

Buchanan, N. T., Perez, M., Prinstein, M. J., & Thurston, I. B. (2021). Upending Racism in Psychological Science: Strategies to Change How Science Is Conducted, Reported, Reviewed, and Disseminated. *American Psychologist, 76*, 1097–1112. https://doi.org/10.1037/amp0000905

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children' s mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273–293. https://doi.org/10.1207/S15326942DN1903_3

Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage, 46*, 338–342. https://doi.org/10.1016/j.neuroimage.2009.01.064

Campbell, C., Landry, O., Russo, N., Flores, H., Jacques, S., & Burack, J. A. (2013). Cognitive flexibility among individuals with Down syndrome: Assessing the influence of verbal and nonverbal abilities. *American Journal on Intellectual and Developmental Disabilities, 118*(3), 193–200. https://doi.org/10.1352/1944-7558-118.3.193

Canfield, R. L., Gendle, M. H., & Cory-Slechta, D. A. (2004). Impaired Neuropsychological Functioning in Lead-Exposed Children. *Developmental Neuropsychology, 26*, 513–540.

Carey, S. (2009). *The Origins of Concepts*. Oxford University Press.

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science, 11*, 339–353.

Chiappedi, M., Mensi, M., Antonaci, E., Zavani, E., Tronconi, L., Termine, C. & Balottin, U. (2018). Intellectual Profile of Adolescents with Headache: A Case– Control Study Using the WISC-IV. *Frontiers in Neurology. 9*. https://doi.org/10.3389/fneur.2018.00128

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99*, 45–77.

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

Crone, E. A., & Van der Molen, M. W. (2004). Developmental Changes in Real Life Decision Making: Performance on a Gambling Task Previously Shown to Depend on the Ventromedial Prefrontal Cortex. *Developmental Neuropsychology, 25*, 251–279.

Cruishank, J. (2003). *Critical Realism: The Difference It Makes*. Routledge.

Darlington, R. B. (1971). Another look at "cultural fairness," *Journal of Educational Measurement, 8*, 71–82.

Dauvier, B., Bailleux, C., & Perret, P. (2014). The Development of Relational Integration During Childhood. *Developmental Psychology, 50*, 1687–1697. https://doi.org/10.1037/a0036655

Delis, D. C., Kaplan, E. & Kramer, J. H. (2001). *The Delis-Kaplan Executive Function System.* Psychological Corporation.

Diamond, A. (1988). Differences between adult and infant cognition: Is the crucial variable presence or absence of language? In L. Weiskrantz (Ed.), *Thought without language* (pp. 337–370). Oxford University Press.

Fernandez, A. L., & Abe, J. (2017). Bias in cross-cultural neuropsychological testing: problems and possible solutions. *Culture and Brain, 6,* 1–35.

Flanagan, O. J. (1981). Psychology, progress, and the problem of reflexivity: A study in the epistemological foundations of psychology. *Journal of the History of the Behavioral Sciences, 17*(3), 375–386. https://doi.org/10.1002/1520-6696(198107)17:3%3C375::aid-jhbs2300170308%3E3.0.co;2-u

Garret, H. E. (1947). N****[word omitted due to offensive outdated language]-White differences in mental ability. *Scientific Monthly, 65,* 329–333.

Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (5th ed., Vol 2, pp. 575– 630). Wiley.

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

Golden, C. J. (1981). The Luria-Nebraska Children's Battery: Theory and formulation. In G.W. Hynd & J. E. Obrzut (Eds.), *Neuropsychological assessment and the school-age child* (pp. 277–302). Grune & Stratto.

Goldstein, S., Naglieri, J. A., Princiotta, D., Otero, T. M. (2014). Introduction: A History of Executive Functioning as a Theoretical and Clinical Construct. In: S., Goldstein, J., Naglieri, (Eds.), *Handbook of Executive Functioning.* Springer. https://doi.org/10.1007/978-1-4614-8106-5_1

Gottesman, I. I. (1968). Biogenetics of race and class. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.) *Social Class, Race and Psychological development*. Holt, Rinehart and Winston Inc.

Greve, K. W., Love, J. M., Dickens Jr., T. J., & Williams, M. (2000). Developmental Changes in California Card Sorting Test Performance. *Archives of Clinical Neuropsychology, 15*, 243–249.

Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., Bundy, D. A., & Sternberg, R. J. (2001). The organization of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development, 25*, 367–378.

Haapasalo, L., & Kadjievich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik, 21*, 139–157.

Hayes, B. K., Fritz, K., & Heit, E. (2013). The relationship between memory and inductive reasoning: Does it develop? *Developmental Psychology, 49*(5), 848–860. https://doi.org/10.1037/a0028891

Heaton, R. K. (1981). *A manual for the Wisconsin Card Sorting Test.* PAR.

Helms, J. E. (1992). Why is there no study of Cultural Equivalence in standardized cognitive Ability Testing? *American Psychologist, 47*, 1083–1101.

Helms, J. E. (2002). A remedy for the Black-White test-score disparity. *American Psychologist, 57*(4), 303–305. https://psycnet.apa.org/doi/10.1037/0003-066X.57.4.303b

Helms-Lorenz, M., & Van de Vijver, F. J. R. (1995). Cognitive Assessment in Education in a Multicultural Society. *European Journal of Psychological Assessment, 11*(3), 158–169.

IBM Corp. (2021). *IBM SPSS Statistics for Windows* (28.0.0.0) [Computer software]. IBM Corp.

Izik Taner, Y., Erdogan Bakar, E., & Oner, O. (2011). Impaired executive functions in paediatric obsessive-compulsive disorder patients. *Acta Neuropsychiatrica, 23*(6), 272–281. https://doi.org/10.1111/j.1601-5215.2011.00562.x

Jacobs, R., Anderson, V., & Harvey, S. (2001). Concept generation test as a measure of conceptual reasoning skills in children: Examination of developmental trends. *Clinical Neuropsychological Assessment, 2*, 101–117.

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology, 20*(3), 573–591. https://doi.org/10.1207/875656401753549807

Johnson, T. P. (2006). Methods and frameworks for crosscultural measurement. *Medical Care, 44*, 17–20.

Kaarakainen, M. T., & Saikkonen, L. (2022). Remark on digital accessibility: educational disparities define digital inclusion from adolescence onwards. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-022-00908-5

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* MIT Press.

Keita, S. O. Y., Kittles, R. A., Royal, C. D. M., Bonney, G. E., Furbert-Harris, P., Dunston, G. M., Rotimi, C. N. (2004). Conceptualizing human variation. *Nature Genetics, 36*(11), 17–20.

Keller, A. E. (2021). *Neuropsychological Test Performance in Anxious Youth: An Examination of the impact of environmental factors and anxiety assessment.* Boston University.

Kharitonova, M., Chien, S., Colunga, E., & Munakata, Y. (2009). More than a matter of getting 'unstuck': Flexible thinkers use more abstract representations than perseverators. *Developmental Science, 12*(4), 662–669. https://doi.org/10.1111/j.1467-7687.2008.00799.x

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimensional-switching task. *Developmental Science, 6*, 449–476.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., Bostrom, A., Sinha, L., Berhel, A., & Widmeyer, M. (2014). NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society, 20*(1), 11–19. https://doi.org/10.1017/S1355617713001094

Kunda, M., Soulieres, I., Rozga, A., & Goel, A. K. (2016). Error patterns on the Raven's Standard Progressive Matrices Test. *Intelligence, 59*, 181–198. https://doi.org/10.1016/j.intell.2016.09.004

Lee, D., Riccio, C .A., & Hynd, G. W. (2004). The Role of Executive Functions in Attention Deficit Hyperactivity Disorder: Testing Predictions from Two Models. *Canadian Journal of School Psychology, 19*(1–2), 167–189. https://doi.org/10.1177/082957350401900109

McGee, C. L., Schonfeld, A. M., Roebuck-Spencer, T. M., Riley, E. P., & Mattson, S .N. (2008). Children with heavy prenatal alcohol exposure demonstrate deficits on multiple measures of concept formation. *Alcoholism: Clinical and Experimental Research, 32*(8), 1388–1397. https://doi.org/10.1111/j.1530-0277.2008.00707.x

McGonigle-Chalmers, M., & Alderson-Day, B. (2010). Free classification as a window on executive functioning in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(7), 844–857. https://doi.org/10.1007/s10803-010-0947-5

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Mischel, W. (2014). *The Marshmallow Task: Understanding Self-Control and How to Master it.* Corgi.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses:

the PRISMA statement. *British Medical Journal (Clinical research ed.), 339*, b2535. https://doi.org/10.1136/bmj.b2535

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology, 40*, 255–265.

Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child Development, 77*, 395–412.

Park, S. (2017). The state of digital inequalities: interplay between social and digital exclusion. In: S. Park, (Ed.) *Digital Capital*, pp. 35–62. Palgrave Macmillan.

Passier M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1,* 349–370.

Pavitt, A. (2017). *Using a Game-like Procedure as a new Test of Problem Solving and Concept Formation in Children.* [Professional Doctoral Dissertation, University of East London]. UEL Repository. https://repository.uel.ac.uk/item/84xw4d.

Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant and Child Development, 11*, 93–105.

Perraudin, S., & Mounoud, P. (2009). Contribution of the priming paradigm to the understanding of the conceptual developmental shift from 5 to 9

years of age. *Developmental Science, 12*(6), 956–977. https://doi.org/10.1111/j.1467-7687.2009.00847.x

Phillips, B. A., Conners, F. A., Merrill, E., & Klinger, M. R. (2014). Rule-based category learning in Down syndrome. *American Journal on Intellectual and Developmental Disabilities, 119*(3), 220–234. https://doi.org/10.1352/1944-7558-119.3.220

Pugh, G. E. (1977). *The Biological Origin of Human Values.* Basic Books.

Raven, J. (1989), The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States. *Journal of Educational Measurement, 26*, 1–16. https://doi.org/10.1111/j.1745-3984.1989.tb00314.x

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford Psychologists Press.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modelling and Analytics, 2*(1), 21–33.

Regard, M., Strauss, E., & Knapp, P. (1982). Children's Production on Verbal and Non-Verbal Fluency Tasks. *Perceptual and Motor Skills, 55*(3), 839–844. https://doi.org/10.2466/pms.1982.55.3.839

Rennie, D. A. C., Bull, R., & Diamond, A. (2004). Executive Functioning in Preschoolers: Reducing the Inhibitory Demands of the Dimensional Change Card Sort Task. *Developmental Neuropsychology, 26*(1), 423–443. https://doi.org/10.1207/s15326942dn2601_4

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 9*3, 346 –362. https://doi.org/10.1037/0022-0663.93.2.346

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, & F. S., Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Salmond, S. S. (2008). Evaluating the Reliability and Validity of Measurement Instruments. *Orthopaedic Nursing 27*(1), 28–30. https://doi.org/10.1097/01.NOR.0000310608.00743.54

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research, 49*, 252–279.

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations Among Conceptual Knowledge, Procedural Knowledge, and Procedural Flexibility in Two Samples Differing in Prior Knowledge. *Developmental Psychology, 47*, 1525–1538

Siegler, R. S. (1991). Developmental Sequences within and between concepts. *Monograph of the Society of Research in Child Development, 46*(2), 1–84.

Smidts, D. P., Jacobs, R., & Anderson, V. (2004). The Object Classification Task for children (OCTC): A measure of Concept Generation and Mental Flexibility in Early Childhood. *Developmental Neuropsychology, 26*, 385–401.

Thorell, L. B. & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A New Rating Instrument for Parents and Teachers. *Developmental Neuropsychology, 33*(4), 536–552.

Van Bers, B. M. C. W., Visser, I., & Raijmakers, M. (2014). Preschoolers can form abstract rule representations regardless of cognitive flexibility. *Journal of Experimental Child Psychology, 124*, 50–66. https://doi.org/10.1016/j.jecp.2014.01.017

Van der Sluis, S., de Jong, P. F., & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449. https://doi.org/10.1016/j.intell.2006.09.001

Van de Vijver, F. J. R., & Poortinga, Y. H. (1992). Testing in culturally heterogeneous populations: When are cultural loadings undesirable? *European Journal of Psychological Assessment, 8*, 17–24.

Vodegel Matzen, L. (1994). *Performance on Raven's Progressive Matrices: What makes a difference?* Unpublished thesis, University of Amsterdam, The Netherlands.

Wechsler, D. (1974). *Wechsler intelligence scale for children - revised.* Psychological Corporation.

Wechsler, D. (2003). *Wechsler intelligence scale for children (WISC-IV)* (4th ed.). The Psychological Corporation

Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). Pearson.

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

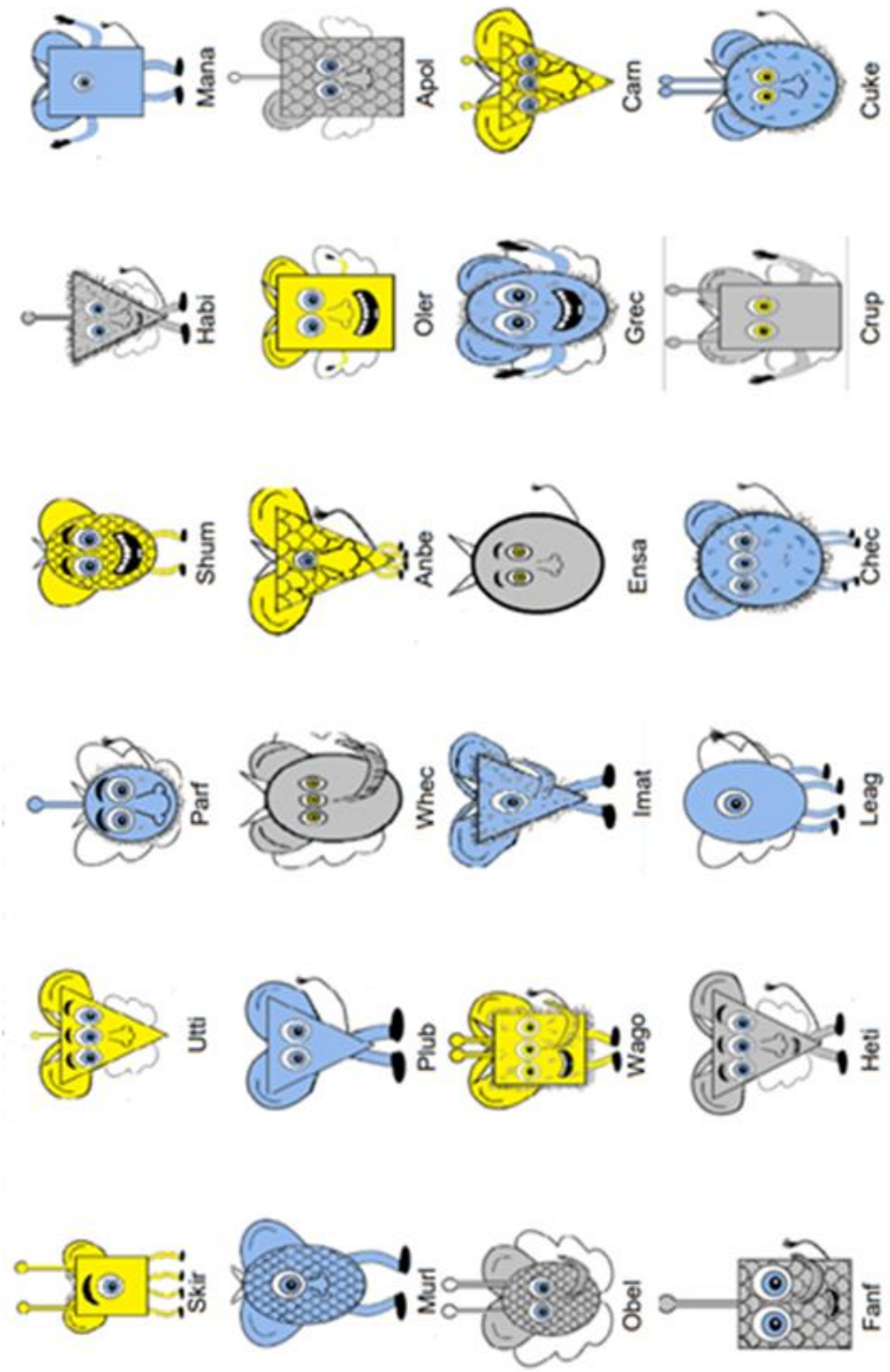
Yeates, K. O., & Donders, J. (2005). The WISC-IV and Neuropsychological Assessment. In A. Prifitera, D .H. Saklofske & L. G. Weiss (Eds.), *WISC-IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives* (pp 415-434). Elsevier Academic Press.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development; 11*, 37–63

Zelazo, P. D., & Müller, U. (2002). The Balance Beam in the Balance: Reflections on Rules, Relational Complexity, and Developmental Processes. *Journal of Experimental Child Psychology, 81*(4), 458–465. https://doi.org/10.1006/jecp.2002.2667

6. APPENDICES

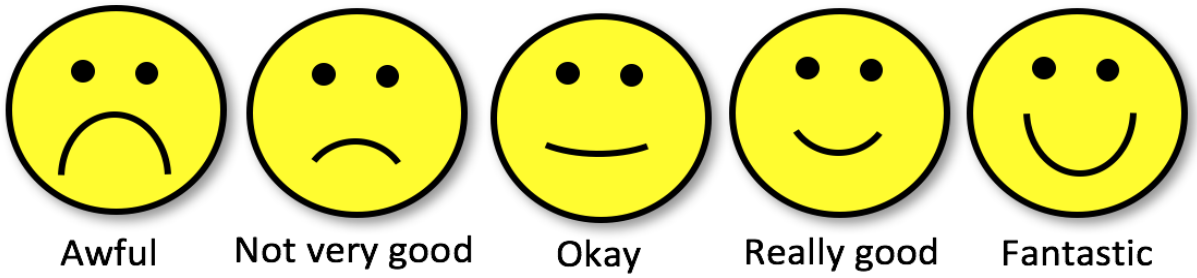
6.1 Original Alien Game set by Pavitt (2017)



6.2 Enjoyability Rating Scale

How enjoyable was the task today?

Tick the face that shows how you felt:



6.3. Alien Game Standardised Instructions

Initial instructions

> *"Now we are going to play the Alien Game. These are the aliens. The game is that I am thinking of one of these aliens and you need to find which one I am thinking of by asking me yes or no questions. This means you can only ask me questions I can answer with a yes or a no. You can ask me up to 10 questions to find the alien, and you need to find it as quickly as you can."*

For the cards condition the additional information was given, after a child had asked an elimination question (e.g. "Is it number 4?", or "Does it have a nose?"):

> *"I will now flip alien Number 4 over."*
>
> Or
>
> *"I will now flip the aliens that have/don't have a nose over."*

If the child struggled with the yes/no format, the following feedback was given.

> "*I cannot answer that question because it is not a yes or no question. Can you change your question so I can answer it with a yes or a no?"*

Sometimes children made statements instead of asking a question (e.g. "I think it has teeth."). In this case, the researcher asked a clarifying question in the following way:

> *"Did you mean to ask if the alien has teeth?"*

If the child asked a yes or no question based on something unrelated to the aliens' characteristics (e.g. "Does it like to play board games?"), the researcher said:

> *"I can't tell by looking at them. Can you think of another question?"*

If the child found the alien within 10 questions, the researcher said:

> "*Yes that is the alien I was thinking of! Now let's try again. Remember that you need to find the next alien I am thinking off as quickly as you can by asking me yes or no questions. You can ask up to ten questions."*

If the child did not find the alien within 10 questions, the researcher said:

> *"It's OK. Let's try again. Remember that you need to find the next alien I am thinking off as quickly as you can by asking me yes or no questions. You can ask up to ten questions.*

If the child asked "Does the alien have feet?" or "Does the alien walk?", that was interpreted the same as "Does the alien have legs?".

6.4 The counterbalancing schedule of the four trials of the Alien Game

If the name of the condition appears in brackets, it means the participant did not complete that trial.

Participant order	Trial 1	Trial 2	Trial 3	Trial 4
1	Board	Board	Cards	(Cards)
2	Board	Cards	Board	(Cards)
3	Cards	Cards	(Board)	(Board)
4	Board	Board	Cards	(Cards)
5	Board	Cards	(Board)	(Cards)
6	Cards	Board	Cards	Board
7	Cards	Cards	Board	Board
8	Board	Cards	Board	Cards
9	Cards	Board	Cards	Board
10	Board	Board	Cards	Cards
11	Cards	Cards	Board	Board
12	Board	Cards	Board	Cards
13	Cards	Board	Cards	Board
14	Board	Cards	Cards	Board

6.5 Ethics Approval Application and Letter of Approval

6.5.1 Ethics Application

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 Procedure as a Test of Executive Functions in Children**
3.2	**Summary of study background and aims (using lay language):**	Concept formation is an executive function and can be understood as the ability to identify relationships between objects or events. It is important to understand executive functioning in children, as these abilities have been found to predict school attainment better than IQ (Blair & Razza, 2007). However, most available tests of executive functioning were designed for adults, and are culturally specific. Pavitt (2017) created 'The Alien Game' based on the format of the children's game "Guess Who?" as a more culture fair test of concept formation for children. Pavitt (2017) ran a pilot study to test this approach, and from her results, she identified several areas for improvement. The current study proposes to further develop The Alien Game in the following ways: (a)to improve and refine the materials used in the game; (b) to design a scoring system based on an established measure of concept formation, modified with Pavitt's (2017) suggestions; and (c) to gather richer data on the feasibility of this game as a culturally fair test of concept formation. This study aims to recruit 60-90 children aged 6-11 years. This study will have a cross-sectional correlational design. In addition to playing the game, participants will be asked to complete two existing measures of concept formation, to address concurrent validity. To compare to real-world executive functioning (criterion validity), class teachers will be asked to complete a questionnaire rating (CHEXI) of

		the child's executive function. We will also address associations between participant demographic data (age, sex and English language facility) and test performance.
3.3	**Research question(s):**	Can a culturally fair test of concept formation be produced that will be engaging to children? Can normative performance characteristics, such as scores and common patterns of responding, which identify normal variation of concept formation be established? Do children engage well with the Alien Game as a measure of concept formation? To what extend will the Alien Game scores of 6 – 11 year-olds correlate with other measures of concept formation, e.g. WISC-IV Matrix Reasoning and Similarities scores? Will participant demographics play a role in moderating the correlation between Alien Game scores and WISC-V Matrix reasoning and Similarities scores?
3.4	**Research design:**	**This study will have a cross-sectional correlational design. Depending on data distributions, parametric or non-parametric procedures (e.g., correlation coefficients, followed up with GLM or regression procedures) will be used to analyse the data and address which variables make unique contributions to test performance. Qualitative data will be used to consider how participants approach the task. Qualitative data will be gathered to understand the quality of the participants responses and strategies, and to determine engagement. Qualitative feedback will be used to determine task enjoyment.**
3.5	**Participants:** Include all relevant information including inclusion and exclusion criteria	**Participants will be recruited from mainstream primary schools in the London region. As this test aims to be culturally fair, we aim to recruit a sample from a range of backgrounds and abilities. The study will aim to accommodate all needs, and not exclude any participants. Participants will be required to have sufficient English abilities or have**

		an interpreter present to consent to participate. Children with sensory and/or motor function impairments will be included where possible if they volunteer.
3.6	**Recruitment strategy:** Provide as much detail as possible and include a backup plan if relevant	**Recruitment of children will be completed through primary schools. Primary schools within London will be contacted via email with details of the study and a poster (see Appendix I) inviting them to take part. A telephone call will be arranged to discuss the details including access to the school and data collection process. We will email the school with all necessary documents and ask them to print information sheets (accessible format for the children) and consent forms for the children and their guardian to read in order to decide whether to take part. Schools will be given the option of using opt-in or opt-out procedure to gain parental consent. Consent will also be gained by the school via the in Loco Parentis form (appendix B). Parents are asked to contact us via email if they have any questions about the study. We will introduce and discuss the study with the child and seek consent or assent as appropriate. Children and parents will be told that they can withdraw their data from the study until the end of the January 2023 if the child / guardian / school change their mind and can stop the study at any point during data collection. Recruitment plan B: To reach out to friends and family who have children within the age range of 6-11, and to recruit via word-of-mouth using the poster (appendix B).**
3.7	**Measures, materials or equipment:** 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**	**YES** ☒	

	attached a GRA form as an appendix:			
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 survey), regardless of the location of the researcher or the participants.	**YES** ☐		
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:**	Alexandros Bardis: 001584640901	
	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**	**YES** ☒

	and feasibility of this research proposal with my supervisor:	
8.2	**Student's name:** (Typed name acts as a signature)	**Alexandros Bardis, Emily Hay and Pinar Marasli**
8.3	**Student's number:**	**U2075206; U2075197; U2075213**
8.4	**Date:**	**20/06/2022**
Supervisor's declaration of support is given upon their electronic submission of the application		

Appendix A: Organisation Invitation Letter

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 January[1] 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

[1] Note for Ethics Application: This date may change depending on when data collection begins- Participants will be given at least six weeks to withdraw their data

Appendix B: Head Teacher's *Loco Parentis* Form

UNIVERSITY OF EAST LONDON

Using a game-like task as an assessment of concept formation in children

Head Teacher's *Loco Parentis* Form

The study (title as above) has been fully explained to me. I have been given the opportunity to review the materials and ask questions.	☐
The parents/guardians of the children who will be invited to participate in this study have been sent a letter home on *[date]* to inform them about the research.	☐
Parents/guardians have been advised that they have a certain period of time (2 weeks) to withdraw (or 'opt-out') their child from participating in the study if they do not wish for them to take part.	☐
I, as the head teacher of the school, am willing to act *in loco parentis* in giving my consent. for the children (whose parents/guardians do not contact me) to participate in the study if they wish to.	☐

Name of head teacher (BLOCK CAPITALS)

……

Name of school (BLOCK CAPITALS)

……

Signature of head teacher

……

Date

……

Researcher’s Name (BLOCK CAPITALS)

……

Researcher’s Signature

……

Date

……

Appendix C: 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 January[2] 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

[2] Note for Ethics Application: This date may change depending on when data collection begins- Participants will be given at least six weeks to withdraw their 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)

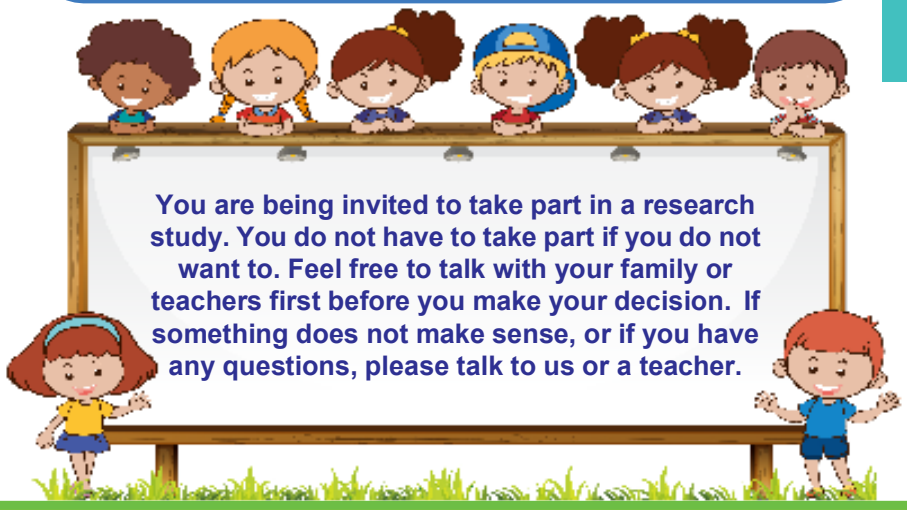
Thank you for taking the time to read this information sheet

Appendix D: Participant 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 E: Parental Consent Forms

UNIVERSITY OF EAST LONDON

Consent to participate in a research study

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

I confirm that I have read the information sheet for the above study and that I have been given a copy to keep. ☐

I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

I understand that my child's participation in the study is voluntary and that I may withdraw them at any time, without explanation or disadvantage ☐

I understand that if I withdraw during the study, my child's data will not be used. ☐

I understand that I have until the end of January[3] 2022 to withdraw my child's data From the study. ☐

I understand that 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, to which I give my permission. ☐

[3] Note for Ethics Application: This date may change depending on when data collection begins- Participants will be given at least six weeks to withdraw their data

It has been explained to me what will happen to the data once the research has been completed. ☐

I understand that anonymised data may be used in material such as conference presentations, reports, articles in academic journals resulting from the study and that these will not personally identify me. ☐

I would like to receive a summary of the research findings once the study has been completed and am willing to provide contact details for this to be sent to. ☐

I agree for my child to take part in the above study. ☐

Participant's Name (BLOCK CAPITALS)

……

Signature of Legal Guardian

……

Researcher's Name (BLOCK CAPITALS)

……

Researcher's Signature

……

Date

……

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 F: Participant Consent Form [4]

Consent to participate in a research study

University of East London

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):	Researcher's Name (BLOCK CAPITALS):
Your Signature:	Researcher's Signature:
Date:	Date:

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

[4] Note for Ethics Application: This date may change depending on when data collection begins- Participants will be given at least six weeks to withdraw their data

Appendix G: Participant Debrief Sheet template – parents

PARTICIPANT DEBRIEF SHEET

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

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

How will my data be managed?

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

What will happen to the results of the research?

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

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

What if I been adversely affected by taking part?

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

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

If you have any questions or concerns about how the research has been conducted, please contact our research supervisor, Dr Matthew Jones Chesters, School of Psychology, University of East London, Water Lane, London E15 4LZ,
Email: m.h.jones-chesters@uel.ac.uk.

or

Chair of School Research Ethics Committee: Dr Trishna Patel, School of Psychology, University of East London, Water Lane, London E15 4LZ.
(Email: t.patel@uel.ac.uk)

Thank you for taking part in our study

Appendix H: Participant Debrief Sheet template – child

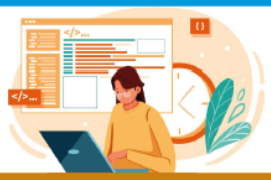
PARTICIPANT DEBRIEF SHEET

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 I: Study Advertisement

University of East London
SCHOOL OF PSYCHOLOGY

THE ALIEN GAME

A new cognitive assessment for children

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

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

CONTACT DETAILS:

alien.game@uel.ac.uk

Our Team:


Alex Bardis


Pinar Marasli


Emily Hay


Meabh Foley

PARTICIPANTS

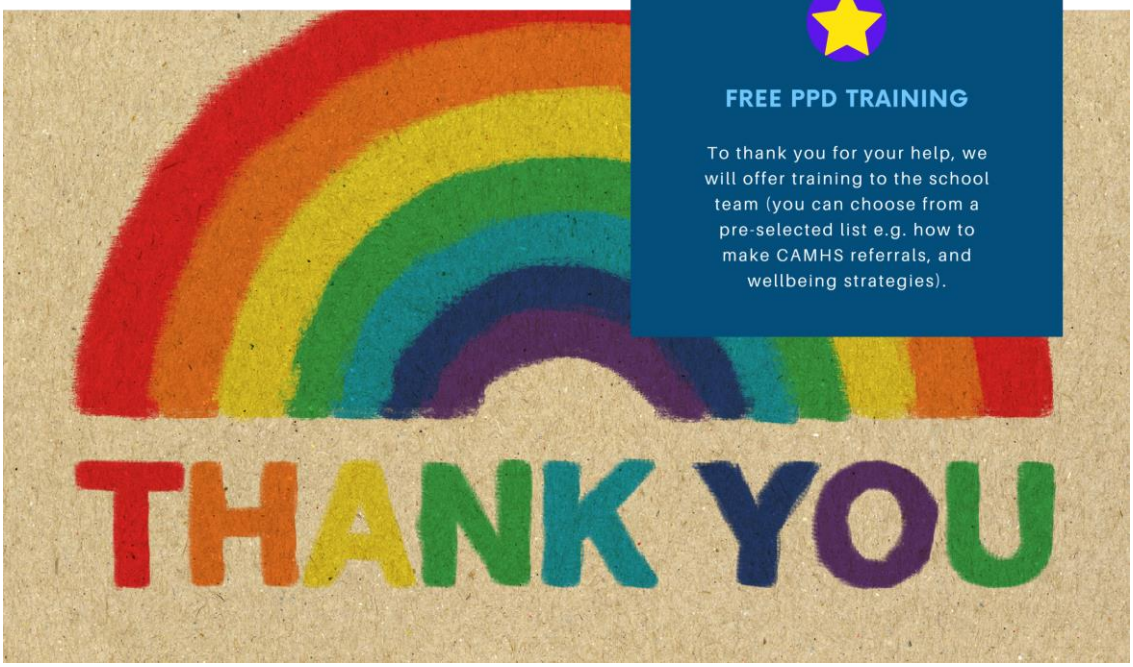
Pupils from Years 1 to Year 6 can participate with parent's consent.

STUDY

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

FREE PPD TRAINING

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



Appendix K: General Risk Assessment Form template

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 Autumn And Winter school term, from October 2022 to January 2023. We hope to assess 4 young people in a day each, which means we will need to spend around 8 days in the school, which will be spread out across the two terms, depending on the school's availability. We plan to meet with each young person individually to complete a battery of neuropsychological assessments, a newly developed game assessing executive functioning and a demographic questionnaire. Teachers will also be asked to complete a questionnaire about the participants behaviour. We plan to complete each session with a young person within an hour. When in the school and meeting with the young person, w will wear a mask at all times and regularly wash and sanitise our hands and any equipment. If possible, we will also request a room with ventilation and the ability to social distance from one another. The resources we will be using are neuropsychological tests, questionnaires and the newly developed game, all of which will be provided by ourselves. We will also provide the school with a copy of our DBS certificates.

Overview of FIELD TRIP or EVENT:
As above

Guide to risk ratings:

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

Hazards attached to the activity

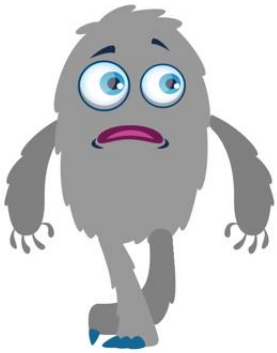
Hazards identified	Who is at risk?	Existing Controls	Likelihood	Severity	Residual Risk Rating (Likelihood x Severity)	Additional control measures required (if any)	Final risk rating

Obstruction of safe exit routes in event of fire or other emergency, due to blocking of doors/thoroughfare/ fire exit routes with tables, chairs or banners.	Staff Students Researcher	On day, researchers will make sure they are aware of where the fire exits are in relation to the location/room used and make sure tables and chairs do not obstruct exits/entrances or routes.	1	2	2	Ensure placement of objects is monitored throughout the day.	2
Slip or trip hazard due to promotional literature or freebies, or rubbish, being dropped on the floor.	Staff, Students, Researcher	Be vigilant on the day to make sure that belongings do not get left on the floor, ensuring anything that is dropped is picked up immediately and ensuring electrical equipment, such as a laptop charger, is in an appropriate place and not a trip hazard. Ensuring bins and cleaning equipment such as paper towels are available.	2	1	2	Ensure this is monitored throughout the day.	2

Infection of covid-19	Ourselves and students whom participate	Wearing a face mask at all times, social distancing where possible, being in a ventilated room, washing and sanitising hands and equipment regularly. Any students who display symptoms or test positive for covid will not participate for at least 14 days, likewise, if a researcher displays symptoms or test positive for covid, that individual will not visit schools for at least 14 days. we also have received both doses of our covid-19 vaccine and booster.	2	2	4		4

Review Date

Appendix M: Example of Alien Game Materials Developed by Researchers



Appendix N: Measures

Demographic Questionnaire

Date of birth: _____

Gender Identity: _____

Ethnicity: _____

Country of birth: _____

Have you always lived in the UK? _____

First language: _____

Main language spoken at home: _____

Parental job title (if known): _____

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

_____
_____
_____

Any difficulties with vision or hearing?

_____
_____

Observation Sheet

Date of testing: _____ Participant ID: _____

Questions asked by participant:

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

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

_____
_____
_____
_____
_____

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

_____
_____
_____
_____
_____

CHILDHOOD EXECUTIVE FUNCTIONING INVENTORY (CHEXI) FOR PARENTS AND TEACHERS

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

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

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

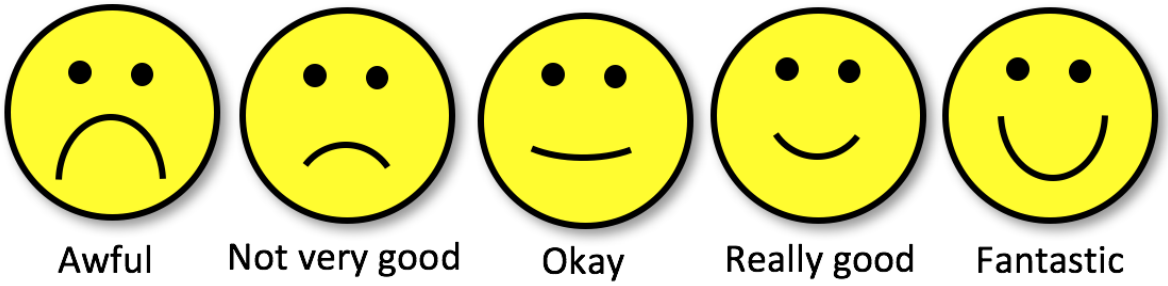
Definitely not true	Not true	Partially true	True	Definitely true
1	2	3	4	5

15. In order to be able to concentrate, he/she 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. For example, he/she needs to jump a couple of 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 he/she is also 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 he/she is doing something else	1	2	3	4	5

Likert Scale Task Enjoyability Measure

How enjoyable was the task today?

Tick the box next to how you felt



Office use only
Child's code:_____

6.5.2 Letter of Ethics Approval

University of East London

School of Psychology Ethics Committee

NOTICE OF ETHICS REVIEW DECISION LETTER

For research involving human participants
BSc/MSc/MA/Professional Doctorates in Clinical, Counselling and Educational Psychology

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

Details	
Reviewer:	**Fiorentina Sterkaj**
Supervisor:	**Matthew Jones Chesters**
Student:	**Emily Hay, Alexandros Bardis, Pinar Marasli**
Course:	**Prof Doc Clinical Psychology**
Title of proposed study:	**Using a Game-Like Procedure as a Test of Executive Functions 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 <u>BEFORE</u> THE RESEARCH COMMENCES	In this circumstance, the student must confirm with their supervisor that all minor amendments have been made **<u>before</u>** the research commences. Students are to do this by filling in the confirmation box at the end of this form once all amendments have been attended to and emailing a copy of this decision notice to the supervisor. The

	supervisor will then forward the student's confirmation to the School for its records. **Minor amendments guidance:** typically involve clarifying/amending information presented to participants (e.g., in the PIS, instructions), further detailing of how data will be securely handled/stored, and/or ensuring consistency in information presented across materials.
NOT APPROVED - MAJOR AMENDMENTS AND RE-SUBMISSION REQUIRED	In this circumstance, a revised ethics application **must** be submitted and approved **before** any research takes place. The revised application will be reviewed by the same reviewer. If in doubt, students should ask their supervisor for support in revising their ethics application. **Major amendments guidance:** typically insufficient information has been provided, insufficient consideration given to several key aspects, there are serious concerns regarding any aspect of the project, and/or serious concerns in the candidate's ability to ethically, safely and sensitively execute the study.

Decision on the above-named proposed research study

Please indicate the decision:	Please select your decision

Minor amendments

Please clearly detail the amendments the student is required to make

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

Major amendments

Please clearly detail the amendments the student is required to make

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

Reviewer's signature	
Reviewer: (Typed name to act as signature)	**Dr Fiorentina Sterkaj**
Date:	**27/10/2022**

This reviewer has assessed the ethics application for the named research study on behalf of the School of Psychology Ethics Committee
RESEARCHER PLEASE NOTE For the researcher and participants involved in the above-named study to be covered by UEL's Insurance, prior ethics approval from the School of Psychology (acting on behalf of the UEL Ethics Committee), and 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)	**Alexandros Bardis**
Student number:	**U2075197**
Date:	**09/11/2022**
Please submit a copy of this decision letter to your supervisor with this box completed if minor amendments to your ethics application are required	

6.6 Additional statistics tables

Table 10

Skewness, Kurtosis and Shapiro-Wilk significance values for the Alien Game IAS, AS, WAS scores, TNQs, CQs, PCQs, RedQs, and RepQs in the Cards and Board condition.

		Skewness Z-score	Kurtosis Z-score	Shapiro-Wilk Sig. (2-sided)
Cards	IAS	-0.18	-1.74	**0.01**
	AS	-0.18	-1.64	**0.04**
	WAS	0.03	-2.01	**0.01**
	TNQs	0.27	-1.52	0.12
	CQs	-0.46	-1.86	**0.00**
	PCQs	0.92	-1.02	**<.001**
	RedQs	**1.57**	1.75	**<.001**
	RepQs	**3.61**	**13.00**	**<.001**
Board	IAS	-0.76	-0.39	**0.01**
	AS	-0.81	-0.30	0.09
	WAS	**1.76**	1.62	**<.001**
	TNQs	**-1.39**	0.40	**<.001**
	CQs	-0.61	-0.45	0.18
	PCQs	**1.66**	1.42	**<.001**
	RedQs	-0.03	-1.80	**0.03**
	RepQs	**2.60**	**7.64**	**<.001**

Note. Values indicating non-normality are in **bold**. IAS= Initial Abstraction Score, AS=Abstraction Score, WAS= Weighted Achievement Score. TNQs=Total Number of Questions, CQs= Number of Conceptual Questions, PCQs= Number of Pseudo-Conceptual Questions, RedQs= Number of Redundant Questions, RepQs= Number of Repetitive Questions. N=13.

Table 11

Spearman's Rank Correlation coefficients for the relationships between Board IAS, AS and WAS scores.

	IAS	AS	WAS
Board IAS	1.00		
Board AS	**.666** (.013)	1.00	
Board WAS	.286 (.344)	.147 (.631)	1.00

Note. Moderate to large effect sizes are in **bold**. IAS= Initial Abstraction Score, AS=Abstraction Score, WAS= Weighted Achievement Score. N=13. *P* values are 2-tailed and are presented in brackets.

Table 12

Spearman's Rank Correlation coefficients for the relationships between Cards IAS, AS and WAS scores.

	IAS	AS	WAS
Cards IAS	1.00		
Cards AS	**.806** (<.001)	1.00	
Cards WAS	**.605** (.029)	**.431** (.141)	1.00

Note. Moderate to large effect sizes are in **bold**. IAS= Initial Abstraction Score, AS=Abstraction Score, WAS= Weighted Achievement Score. N=13. *P* values are 2-tailed and are presented in brackets.

Table 13

Spearman's Rank Correlation coefficients for the relationships between Cards and Board IAS, AS and WAS scores with Age, WISC-IV Similarities and Matrix Reasoning Raw Scores and CHEXI Working Memory, and Inhibition Raw Scores.

		WISC-IV		CHEXI	
	Age (in months)	Similarities	Matrix Reasoning	Working Memory	Inhibition
Cards IAS	.051 (.868)	**.302** (.315)	.293 (.331)	-.158 (.607)	.014 (.963)
Cards AS	-.202 (.508)	**.515** (.072)	**.364** (.222)	**-.306** (.309)	-.144 (.639)
Cards WAS	-.192 (.529)	.157 (.609)	.242 (.425)	-.276 (.361)	-.074 (.810)
Board IAS	**-.587** (.035)	**.605** (.028)	**.697** (.008)	**-.647** (.017)	**-.387** (.191)
Board AS	**-.373** (.209)	**.564** (.045)	**.412** (.162)	-.186 (.543)	-.138 (.653)
Board WAS	.216 (.478)	.011 (.971)	-.030 (.922)	-.114 (.711)	-.167 (.585)

Note. Moderate to large effect sizes are in **bold**. IAS= Initial Abstraction Score, AS=Abstraction Score, WAS= Weighted Achievement Score. N=13. *P* values are 2-tailed and are presented in brackets.

Table 14

Spearman's Rank Correlation coefficients for WISC-IV Similarities and Matrix Reasoning Raw scores, CHEXI Working Memory and Inhibition Raw scores and age of participants.

		Similarities	Matrix Reasoning	Working Memory	Inhibition	Age
WISC-IV	Similarities	1.00				
	Matrix Reasoning	.252 (.406)	1.00			
CHEXI	Working Memory	**-.783** (.002)	**-.417** (.157)	1.00		
	Inhibition	**-.661** (.014)	.214 (.483)	**.661** (.014)	1.00	
Age	Age in months	**-.620** (.024)	**-.536** (.059)	**.582** (.037)	**.410** (.165)	1.00

Note. Moderate to large effect sizes are in **bold**. N=13. *P* values are 2-tailed and are presented in brackets.

Table 15

Spearman's Rank Correlation coefficients for the relationships between Cards TNQs, CQs, PCQs, RedQs, and RepQs.

	TNQs	CQs	PCQs	RedQs	RepQs
Total Number of Questions	1.00				
Conceptual Questions	**-.307** (.308)	1.00			
Pseudo-Conceptual Questions	**.530** (.063)	**-.890** (<.001)	1.00		
Redundant Questions	**.167** (.586)	**.664** (.013)	**-.671** (.012)	1.00	
Repetitive Questions	.196 (.521)	.081 (.792)	-.220 (.470)	**.381** (.200)	1.00

Note. Moderate to large effect sizes are in **bold**. TNQs=Total Number of Questions, CQs= Number of Conceptual Questions, PCQs= Number of Pseudo-Conceptual Questions, RedQs= Number of Redundant Questions, RepQs= Number of Repetitive Questions. N=13. *P* values are 2-tailed and are presented in brackets.

Table 16

Spearman's Rank Correlation coefficients for the relationships between Board TNQs, CQs, PCQs, RedQs, and RepQs.

	TNQs	CQs	PCQs	RedQs	RepQs
Total Number of Questions	1.00				
Conceptual Questions	-.019 (.950)	1.00			
Pseudo-Conceptual Questions	.011 (.972)	**-.754** (.003)	1.00		
Redundant Questions	.293 (.331)	**.518** (.070)	**-.751** (.003)	1.00	
Repetitive Questions	**.352** (.239)	**-.330** (.271)	.020 (.948)	-.074 (.811)	1.00

Note. Moderate to large effect sizes are in **bold**. TNQs=Total Number of Questions, CQs= Number of Conceptual Questions, PCQs= Number of Pseudo-Conceptual Questions, RedQs= Number of Redundant Questions, RepQs= Number of Repetitive Questions. N=13. *P* values are 2-tailed and are presented in brackets.

Table 17

Spearman's Rank Correlation coefficients for the relationships between Card TNQs, CQs, PCQs, RedQs, and RepQs with Age, WISC-IV Similarities and Matrix Reasoning Raw Scores and CHEXI Working Memory, and Inhibition Raw Scores.

			WISC-IV		CHEXI	
		Age (in months)	Similarities	Matrix Reasoning	Working Memory	Inhibition
Cards	Total Number of questions	.073 (.814)	-.209 (.494)	-.172 (.574)	**.407** (.168)	.133 (.666)
	Conceptual Questions	-.188 (.540)	**.618** (.024)	.296 (.326)	**-.311** (.300)	**-.308** (.306)
	Pseudo-Conceptual Questions	**.304** (.312)	**-.705** (.007)	**-.389** (.189)	**.526** (.065)	**.431** (.141)
	Redundant Questions	-.262 (.387)	**.518** (.070)	**.349** (.242)	-.136 (.658)	-.131 (.670)
	Repetitive Questions	**-.309** (.305)	**.464** (.110)	.039 (.899)	**-.465** (.110)	**-.473** (.103)
Board	Total Number of questions	-.147 (.631)	-.209 (.494)	-.172 (.574)	**.407** (.168)	.133 (.666)
	Conceptual Questions	-.188 (.539)	**.618** (.024)	.296 (.326)	**-.311** (.300)	**-.308** (.306)
	Pseudo-Conceptual Questions	.291 (.334)	**-.705** (.007)	**-.389** (.189)	**.526** (.065)	**.431** (.141)
	Redundant Questions	-.262 (.386)	**.518** (.070)	**.349** (.242)	-.136 (.658)	-.131 (.670)
	Repetitive Questions	.099 (.748)	**.464** (.110)	.039 (.899)	**-.465** (.110)	**-.473** (.103)

Note. Moderate to large effect sizes are in **bold**. N=13. *P* values are 2-tailed and are presented in brackets.