Using a game-like task as an assessment of emotion recognition in children.

Méabh Foley

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

Facial emotion recognition, the ability to interpret facial expression of emotion, is a core sociocognitive ability which supports successful social interactions and the development of more complex skills in social cognition such as empathy. Deficits in this area can have a negative impact upon children's performance socially and in education. Despite the importance of this skill most tests of emotion recognition were designed for adults without adaption for children's developmental stage.

This study is an exploratory, first phase development of a new test of emotion recognition for children; *the Alien Quiz*. This test includes two novel additions to emotion recognition assessment for children: a game-like format in the Guess the Alien subtest, and culturally neutral emotion stimuli. Twenty-five children from a mainstream primary school took part in the study to understand whether children apply similar strategies in the Alien Quiz and whether the Alien Quiz could be used as a measure of emotion recognition. Preliminary scoring was developed for all subtests for the task, novel methods of scoring were developed for the game-like task, traditional scoring was adapted for the other subtests.

Content analysis of the Guess the Alien subtest indicated that children used similar patterns of responding within the task and relied upon emotion as a means of categorisation. Descriptive statistics showed that children's accuracy in identifying core emotions using novel cartoon stimuli was similar to accuracy reported in previous studies. Children expressed that they enjoyed the game-like task more than another task.

With amendments noted in the limitations of this study, the Alien Quiz has potential to be used as a measure of emotion recognition for children. Continued research is required in order to assess validity, norms, and reliability of the task prior to its use clinically.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the children who participated in this research. I learned so much from you all; thank you for being so enthusiastic and honest in your participation and feedback. I would also like to thank the school that hosted me, in particular the Head Teacher, SENCO, and teachers who supported communication of the research to parents. Thank you for being so accommodating throughout data collection.

I would like to thank my supervisor, Dr Matthew Jones-Chesters, for his support and input over the past two years. I have really appreciated your guidance, and insights, and of course your thoughtful and containing feedback.

I am deeply grateful to my friends and family for endless support and encouragement throughout this course. I'm so appreciative of having such warm and loving people in my life. Special thanks to my wonderful parents and sister for their constant encouragement and belief in me. Finally, Nico, thank you for your abundance of emotional and practical support, and for ensuring that I had much needed breaks and moments of fun.

I am deeply grateful to you all and would not have been able to do this without you.

Table of Contents

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

ADHD Attention Deficit Hyperactivity Disorder

ASD Autism Spectrum Disorder

CAM-C Cambridge Mindreading Face-Voice Battery for Children

CV Coefficient of Variation

DANVA Diagnostic Analysis of Nonverbal Accuracy

DEEP Developmental assessment on an E-Platform

FEEL Facially Expressed Emotion Labelling

FEEST Facial Expressions of Emotion: Stimuli and Tests

FEFA-2 The Frankfurt Test and Training of Facial Affect Recognition

FELT Facial Expression Labelling Task

FEM Facial Emotion Matching

FEN Frequency of Emotion Naming

FEPT Facial Emotion Perception

FER Facial Emotion Recognition

GTA Guess the Alien

NEPSY-II Developmental NEuroPSYchological Assessment-II

SDQ Strengths and Difficulties Questionnaire

UEN Use of Emotion Naming

1. INTRODUCTION

This research is focused upon assessment of emotion recognition in children. This is a developmental sociocognitive ability and is located within the broader area of social cognition. Since the emergence of social neuroscience in the 1990's (Beaudoin & Beauchamp, 2020), researchers have in increasing numbers explored the development of sociocognitive skills throughout the lifespan, and the impact that failure to acquire such skills may have on children as they progress to adulthood (Cotter et al., 2018). Despite the breadth of research ongoing in this area, tests of social cognition for children often repurpose those first designed for adults, without considering their appropriateness for the developmental stage of the child, their ability to engage with the test, and the cultural appropriateness of the test. The current study aims to assess the utility of using novel, culturally neutral, emotion stimuli, and a game-like structure for a test of emotion recognition in children.

The aim of this section is to orientate the reader to the research area. An introduction to social cognition will be provided in order to facilitate the positioning of emotion recognition within this larger research area. The developmental trajectory of facial emotion recognition (FER) and aetiology of the difficulties linked to deficits in FER will be explored in the discussion of emotion recognition. A critical review of two independent literature searches will then be provided: the first exploring the development of tests of FER, the second exploring the development of game-like tests of cognition. Finally, the rational and clinical need for the current study will be explored in relation to gaps in the current literature.

1.1. Social Cognition

The term cognition refers to a number of different processes through which both humans and other living organisms create an understanding of the world around them (Frith, 2008). Successful social interactions require the individual to utilise core cognitive processes such as perception, attention, and memory (Frith, 2008) in combination with abilities more specific to social interactions know as social cognition (Beaudoin & Beauchamp, 2020).

Social cognition may refer to, and encapsulate, a complex array of cognitive abilities which form the basis of perception, interpretation, process, and response of social information and stimulus (Kobiella et al., 2008; Leppänen, 2011; Somerville et al., 2011). Social cognition within experimental and clinical psychology is often considered to be the ability to perceive and prioritize social stimuli (Morrison et al., 2020), correctly identify emotional states, and the capability to determine the thoughts and intentions of others, known as theory of mind (Baron-Cohen, 1991).

Development of sociocognitive functions unfold during the neonatal period and continue to mature across the lifespan. From birth, social interactions are an integral part of much of human behaviour (Beaudoin & Beauchamp, 2020). As children grow, social behaviour becomes more complex (Kristen et al., 2011). The improvement of a young person's competence in recognising, perceiving and interpreting subtle and evolving social information enhances their social aptitude over the lifespan (Beauchamp & Anderson, 2010; Crick & Dodge, 1994).

Core sociocognitive functions include basic abilities such as facial processing, joint attention, and facial emotion recognition alongside more complex processes such as theory of mind, social decision making and moral reasoning (Kilford et al., 2016). Acquisition of basic sociocognitive functions provides the foundation for the development of more complex abilities. An example of this is facial processing as a prerequisite for facial emotion recognition, which in turn supports recognition of complex cognitive, emotional and social states such as 'anger', 'fear' and 'shame' (Beaudoin & Beauchamp, 2020).

1.1.1. Neurobiological Basis of Social Cognition

Social cognitive processing relies on complex cortical and subcortical neural pathways (Cotter et al., 2018). The ability to evaluate emotional stimuli and socially salient information is part of the amygdala network (Beaudoin & Beauchamp, 2020). Social cognitive abilities are also supported by the temporoparietal junction and prefrontal cortex (Forbes & Grafman, 2010; Molenberghs et al., 2016). Although various brain areas are understood to be associated with facial analysis and

identification (Lopatina et al., 2018), it is thought that mirror systems are one mechanism through which FER effectively functions to enhance social cognitive interactions. Mirror systems refer to regions in the brain which activate when we do something or experience something ourselves, but also when we witness the action or experience in someone else (Frith, 2008). Mirror neurons are thought to function in this way to tell us things about the world which are important for our survival. For example, Wicker and colleagues (2003) identified the mirror system for disgust (whereby the act of witnessing another's expression of disgust triggers a feeling of disgust in ourselves) that functions to warn us that there is something that should be avoided. The use of facial expression to suggest avoidance can also be extended to signal that an individual should be avoided. Untrustworthy faces trigger an automatic response in the amygdala providing physiological signals to avoid that person (Winston et al., 2002). Mirroring responses also function to build a positive rapport with others and provide a sense of positive sentiment both with others and also the environment (Frith, 2008).

1.2. Facial Emotion Recognition

FER refers to the ability to interpret facial expression of emotion (Paiva-Silva et al., 2016). The ability to recognise facial emotions supports successful social interactions and interpersonal functioning in social groups (Schultz, 2005). FER allows for a chain of social stimuli to take on a communicative role. For example, if a group expresses 'anger' following their peer's social faux pas, an expression of embarrassment by their peer may elicit an expression of sorrow in the rest of the group and ultimately resolve the conflict (Frith, 2008). Emotion processing, alongside theory of mind are building blocks onto which support the development of empathy (Bird & Viding, 2014) and later, moral reasoning (Garrigan et al., 2016).

An innate human ability to identify and attribute emotional meaning to facial expression was first proposed by Darwin in 1872 (Darwin & Prodger, 1998). Darwin hypothesised that this ability was universal in humans. This hypothesis was later corroborated by Ekman in the 1960s who conducted cross-cultural research on the universality of emotion recognition by presenting to participants photographs of

actors making different emotional expressions (Ekman, 2003). Through his research Ekman established that there is widespread cross-cultural recognition of six basic emotions; ‘anger’, ‘fear’, ‘joy’, ‘sadness’, ‘surprise’, and ‘disgust’ (Ekman & Friesen, 2003). This research led to the development of the Pictures of Facial Affect test of emotion recognition for adults (Ekman & Friesen, 1976) and later the Facial Expressions of Emotion: Stimuli and Tests (FEEST; Young et al., 2002).

1.2.1. Development of Facial Emotion Recognition in Children

An infant’s ability to discriminate between rudimentary facial emotions emerges around 9-12 months through a process of gradual refinement (Leppänen, 2011; Somerville et al., 2011). The ability to accurately label common emotions appears to emerge gradually throughout early to middle childhood. A period of establishment of FER ability occurs between ages six to eight years, ability is consolidated to adult performance at around 14 years old (Kolb et al., 1992). While there have been mixed findings on the developmental trajectory of FER of certain emotions there is a consensus that the emotion ‘joy’ is the first that young children are able to correctly label (Somerville et al., 2011). Lawrence and colleagues (2015) completed a robust study of FER in 478 UK schoolchildren. They found that 92% of six-year-olds tested could accurately identify the emotion ‘happiness’ or ‘joy’. Lawrence and colleagues (2015) established childhood norms for the Pictures of Facial Affect (Ekman & Friesen, 2003) by showing the image of an adult faces expressing the six basic emotions to children of different age ranges. Strengths of this research included recruitment of a large cohort sample, assessment of the ability to recognise all of the six basic emotions, and a wide age range (6-16 years).

Correct identification of the basic emotions ‘disgust’ (Herba et al., 2008; Khawar et al., 2014; Lawrence et al., 2015; Rodger et al., 2015), ‘surprise’ (Lawrence et al., 2015; Rodger et al., 2015) and ‘fear’ (Lawrence et al., 2015; Thomas et al., 2007a) emerge and improve gradually throughout childhood. There have been conflicting findings on the development of FER of ‘anger’ and ‘sadness’; though most studies indicate that recognition of these emotions also improve throughout the period of middle childhood (ages 6-8 years) (Herba et al., 2008; Khawar et al., 2014; Rodger et al., 2015). In contrast Lawrence and colleagues (2015) found no age effects for

‘anger’ and ‘sadness’, with six- and 16-year-old children achieving similar accuracy levels for both emotions (75-79%). Research on development of FER is limited by differences in FER tests used, with many studies using tools originally designed for adults (Khawar et al., 2014; Lawrence et al., 2015).

There have been conflicting findings on whether FER continues to develop beyond adolescence (Lawrence et al., 2015). Some researchers suggest that prior to adolescence children demonstrate similar FER skills to adults (Kolb et al., 1992; Rodger et al., 2015; Tonks et al., 2007a). Khawar and colleagues (2014) found that out of four age ranges (child, adolescent, adult, older adult) children scored the lowest and adults scored the highest on overall scores of FER. It appears that some emotion identification, for example ‘anger’, continues to develop into adulthood. Thomas and colleagues (2007) compared the accuracy of FER in children and adults found that subtle changes continue to occur in FER of ‘anger’ post adolescence, with the recognition of ‘anger’ significantly increased in young adults in comparison to adolescents. Table 1 below contains a detailed breakdown of studies addressing the development of facial emotion recognition by ages and their findings.

It appears that FER, in line with other sociocognitive abilities, develops in a dynamic fashion rather than through well-defined increments (Garrigan et al., 2016). While functional organisation of sociocognitive abilities in the brain is established by late childhood, further development and neurofunctional specialisation can result in improvement of sociocognitive abilities continuing into adulthood (McCormick et al., 2018). One possible neurological explanation for the variation between emotions on the development of FER is that the later development of the prefrontal cortex during adolescence (Nelson & Guyer, 2011; Sowell et al., 2002) influences the interpretation of emotions of ‘anger’ and ‘fear’ (Yurgelun-Todd, 2007). Further research, using tests designed specifically for children, is required in this area to clarify the developmental trajectories of individual emotions.

Table 1

Developmental Trajectory of Facial Emotion Recognition

Age	Infancy	Early childhood	Mid to late Childhood	Adolescence	Young Adulthood
Facial emotion recognition ability	**New-born** Preference for patterns which are face-like (Schultz, 2005) Preference for attractive/smiling faces (Beaudoin & Beauchamp, 2020; Slater et al., 1998) **Three to nine months** Attentional bias to facial stimuli grows (Frank et al., 2009) Emergence of discrimination of basic facial expression of emotion (Kobiella et al., 2008; Leppänen, 2011; Somerville et al., 2011)	Emergence of ability to identify, match and label six basic emotions (Beaudoin & Beauchamp, 2020) **Three Years** Use of one or two broad descriptors for emotions such as 'happy' and 'sad' (Widen & Russell, 2003) **Five to Six Years** 'Joy', 'sadness' and 'anger' facial expressions are consistently recognisable (Kujawa et al., 2014; Lawrence et al., 2015)	**Six to Eight Years** Increased proficiency in labelling more complex emotions such as 'surprise' and 'disgust' (Kujawa et al., 2014; Lawrence et al., 2015) **Eleven Years** By this age FER through matching and labelling of basic emotions is at near adult levels (Chronaki et al., 2015)	Adolescents demonstrate continued refinement of ability to identify 'fear' and 'disgust' into late teenage years (Lawrence et al., 2015).	Young adults have highest FER scores (Khawar et al., 2014), it is possible that further development of prefrontal cortex contributes to increased performance (Nelson & Guyer, 2011; Sowell et al., 2002)
Emotion word naming* (Baron-Cohen et al., 2010)		**Four to Six years** 'Joy' and 'sadness'	**Seven to Eight years** 'Anger', 'surprise', and 'fear' **Eleven Years** 'Disgust'		

*Defined as more than 90% of children able to understand word

1.2.2. Aetiology of Difficulties in FER

Proposed causes of deficits in facial emotion recognition may be broadly divided into four main sources: brain injury, neurodevelopmental disorders, mental health conditions, and social and environmental causes.

1.2.2.1. Brain injury: it is widely established that brain injury in adults sometimes results in impairment of facial emotion recognition (Schmidt et al., 2010). There has been limited research on the impact of brain injury in childhood; however the few studies exploring this area have confirmed that impairment in FER is common among young people with brain injuries in comparison to age matched controls (Schmidt et al., 2010; Snodgrass & Knott, 2006; Tonks et al., 2007b). Deficits in emotion recognition following brain injury have been linked to damage to the frontal lobe region (Snodgrass & Knott, 2006). Given the absence of a more comprehensive literature base investigating this topic, and the lack of specification of which area of the brain damage occurred in other studies (Schmidt et al., 2010; Tonks et al., 2007b), it is possible that damage to other areas of the brain lead to facial emotion recognition deficits. It is also noteworthy that researchers recruited participants with heterogeneous causes of brain damage (for example, traumatic brain damage, strokes and meningitis) when researching this topic, which has further increased the difficulty in establishing damage to which areas of the brain are linked to difficulties in FER.

1.2.2.2. Neurodevelopmental disorders: difficulties in FER have been identified in children with diagnoses of attention deficit hyperactivity disorder (ADHD) (Airdrie et al., 2018; Collin et al., 2013; Maire et al., 2019); and autism spectrum disorder (ASD) (Dawson et al., 2002; Joseph & Tanaka, 2003; Schultz, 2005; Shanok et al., 2019). FER deficits have also been found in children with specific learning disabilities such as dyslexia and dyscalculia (Operto et al., 2020).

Hypoactivation of the fusiform face area, temporoparietal area, and amygdala region in persons with autism may lead to deficits in the recognition of emotionally salient facial information (Schultz, 2005). It has been hypothesised that the amygdala modulates these processes, with post-mortem studies indicating poor neuronal

arborization within the amygdala of those with autism diagnoses (Bauman & Kemper, 2005). Disturbances in the temporal area of the brain were also found among young people with ADHD diagnoses when exposed to a facial emotion recognition task. These were found to improve with treatment of methylphenidate (Williams et al., 2008).

1.2.2.3. *Mental health conditions:* studies suggest that children with bipolar affective disorder (BPAD) (Brotman et al., 2008; McClure et al., 2005); eating disorders (Castro et al., 2010; Zonnevijlle-Bendek et al., 2002); and anxiety (Simonian et al., 2001; Waters et al., 2008) have deficits in FER. Among adults there is evidence for a relationship between low mood and a reduced accuracy for the FER of 'happy' and 'sad' faces (Bourke et al., 2010). There is limited research on the relationship between young people with low mood and FER, though Simcock and colleagues (2020) did not identify a positive correlation between low mood and FER of 'sad' faces. Young people with anxiety disorders have a bias towards faces displaying negative emotions such as 'anger' and 'fear', and are more likely to make errors in identification of positive or neutral emotions (Collin et al., 2013). FER deficits in BPAD indicate a hyperactivation of the amygdala alongside hypoactivation of the prefrontal cortex, a contrast to neurological findings on FER deficits in neurodevelopmental disorders (Collin et al., 2013).

1.2.2.4. *Social and environmental causes:* adverse early life experiences have also been linked to difficulties in FER (Kujawa et al., 2014). Pollak and colleagues (2000) found that early experience of physical abuse and neglect can influence a child's ability to discriminate emotional expression. Furthermore, children who had been physically abused demonstrated a bias for faces that were angry (Pollak et al., 2000). These findings were confirmed by a meta-analysis which indicated a moderate disadvantage in FER of children who have experiences physical abuse compared to those who have not (Wagner et al., 2015). Paine et al., (2021) found that adopted children were poorer at discriminating 'sad' and 'angry' faces than age matched controls. Systematic review of the literature indicated that across various forms of child maltreatment children develop a negative bias towards recognising negative facial emotions (Assed et al., 2020).

1.2.3. Impact of Deficits in FER

Deficits in FER can have an impact a child's psycho-social functioning in several ways. The sub-section below outlines the social, emotional, and behavioural difficulties that are associated with FER deficits.

1.2.3.1. Social difficulties: the ability to adequately interpret expressions of emotions on faces is a key component of positive social functioning. Difficulties recognising facial emotions have a negative effect on building and sustaining relationships (Collin et al., 2013). The failure to acquire this ability can also impact upon the development of higher order components of social cognition such as theory of mind and empathy (Beaudoin & Beauchamp, 2020; Korkmaz, 2011); compounding difficulties in social functioning. Failure to acquire competencies in social cognition can result in poorer mental health and psycho-social wellbeing in adulthood (Henry et al., 2016). Deficits in social cognition may also have an impact on epistemic trust in early interactions of children (Koenig & Harris, 2005), thus preventing children from identifying which information can be trusted and which information should be treated with scepticism. Increased social cognitive ability has been linked to improved peer relationships in children (Andrés-Roqueta et al., 2016). Children who were relationally bullied demonstrated poorer ability to identify negative emotions in faces than control children (Woods et al., 2009). Children who can accurately recognize and respond to the emotions of others are better able to regulate their own emotions and behaviour (Denham et al., 1997).

1.2.3.2*. Behavioural difficulties:* children with difficulties recognising facial emotions are at risk of behavioural problems. Poor FER contributes to children developing internalising behaviours (Castro et al., 2010). Internalising behaviour describes negative behaviours and thoughts towards oneself, and is associated with low mood, and social withdrawal (Eisenberg et al., 2001). A systematic review of the evidence found that FER deficits are associated with externalising behaviours in children, though the causality of this relationship is uncertain (Cooper et al., 2020). Externalising behaviours are negative behaviours directed at others and the external environment (Eisenberg et al., 2001) One possible explanation may be that children with a bias for recognition of 'anger' exhibited poorer behaviour in social situations. (Barth & Bastiani, 1997)

1.2.3.3. *Mental health difficulties:* researchers have found that difficulties in FER in adulthood are related to suicide attempts even when controlling for low mood. Individuals with a prior suicide attempt demonstrated increased errors in FER in comparison to matched non-suicide attempters (Szanto et al., 2012). De la Torre-Luque et al., 2022) hypothesised that FER deficits may be associated with withdrawal from social situations and feelings of loneliness; withdrawal in turn, may influence the pre-existing FER difficulties and confirm existing biases. There is a paucity of research investigating links between low mood and FER in children. Identification of the relationship between FER and psychological distress in adulthood, highlights the importance of detection of this difficulty and amelioration of the skill in childhood.

1.3. Literature Review

The present study identified a gap in the literature for game-like tests of FER. As a result two independent literature reviews were completed for this thesis. The aim of conducting dual literature reviews was to explore the literature on tests of FER, alongside the development of game-like tests of cognition. This information was then synthesised and utilised to provide insights and guidance for the novel test of FER developed in this study.

1.3.1. Literature Review Method

Two independent literature reviews were completed for this thesis.
Literature review A, was comprised of a scoping review of tests of FER in children. Literature review B, involved a further scoping review to identify game-like tests of cognition in children. Both literature reviews used PsycInfo, CINHAL and Academic Search Ultimate databases. Reference lists of identified papers were manually searched in order to retrieve other relevant publications.

1.3.2. Inclusion and Exclusion Criteria

The review of literature was performed in September 2022. Inclusion criteria for both reviews were English language empirical papers and systematic reviews, published in peer-reviewed journals, publication between January 2000 and September 2022.

Studies were only included if the participant group included typically developing children.

Search A “tests of emotion recognition in children” included studies focused on assessment of FER in typically developing children. Exclusion criteria in this literature review were studies that assessed FER in adults only, studies that assessed FER in neurodiverse groups only, and articles that focused on training of FER skills. Search terms used in Search A were (emotion recognition or emotion perception or emotion identification or emotion labelling) AND (children or adolescents or youth or child or teenager) AND (test or assessment or evaluation or measurement or scale or instrument).

Search B “game-like cognitive tests for children” included studies focused on game-like cognitive assessment tasks in typically developing children. Exclusion criteria in this literature review were studies that focused on training of cognitive skills, articles that focused on adult population groups and articles that focused on neurodiverse population groups. Search B used the search strategy (cognition or cognitive development or cognitive ability or cognitive functioning or cognitive skills or executive functioning or social cognition or emotion recognition) AND game* AND (test or examination or exam or assessment) AND (children or adolescents or youth or child or teenager).

1.4. **Review of Literature - Tests of Facial Emotion Recognition**

This section will review contemporary literature on the assessment of FER in children. An overview of methodology of current tests of FER will be provided; these are categorised by type of stimuli used in assessment. A critical review of the included literature will be provided both within discussion of the papers and in the summary of the section. This literature search included tests of FER for adults only if identified papers indicated that tests were used for child populations.

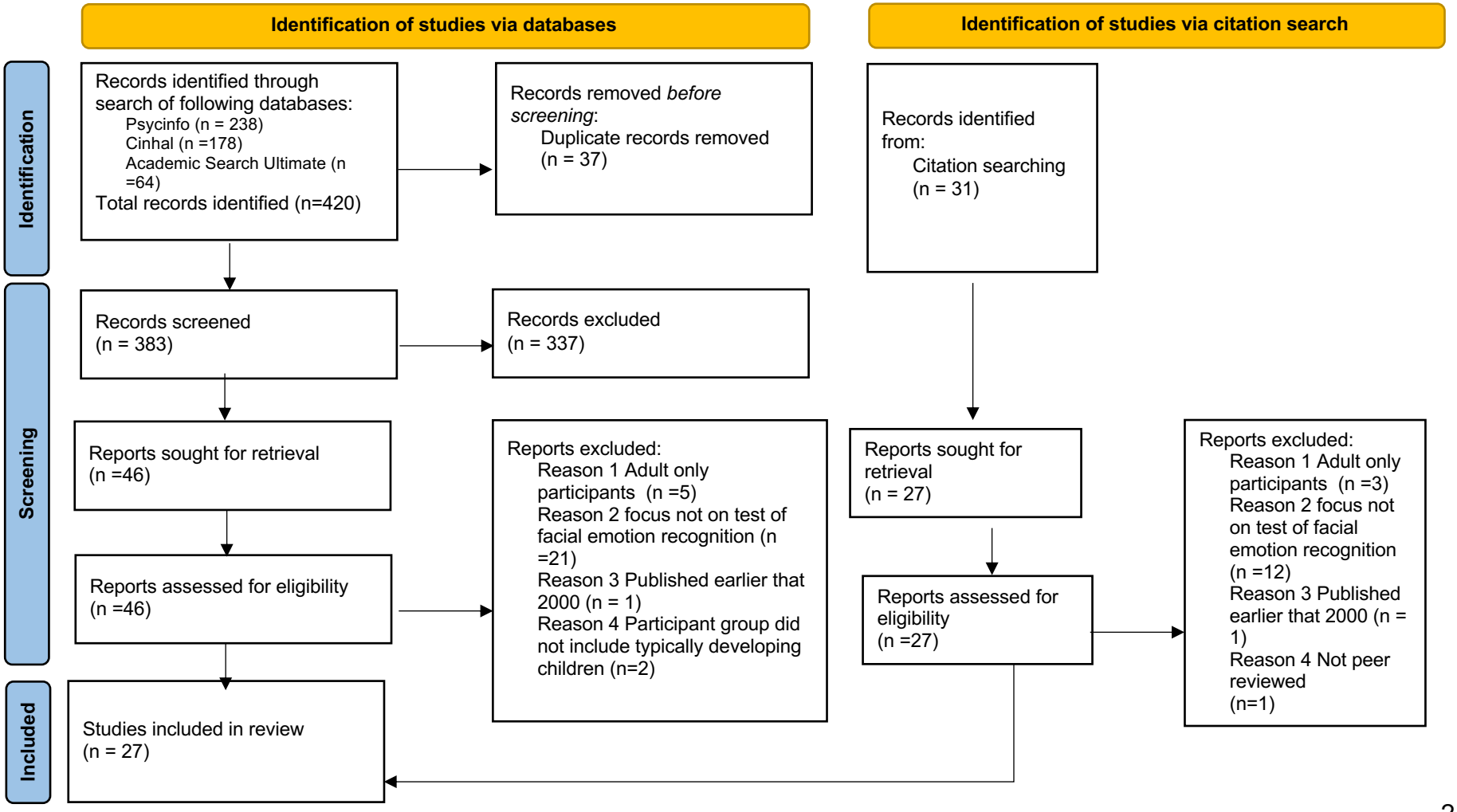
1.4.1. Search Results

Search results from literature search A investigating the literature base for “tests of emotion recognition in children” are displayed in Figure 1 below. The total number of articles identified through search of databases was 420. Duplicate papers were removed, and titles and abstracts were manually screened for eligibility and relevance to research questions. Full text of articles were read for papers which appeared to meet the inclusion criteria. Citation searching of these papers identified an additional potentially germane 30 papers.

A final 27 papers were determined to be eligible for inclusion. The following section of this report provides a qualitative synthesis of material from these papers.

Figure 1

PRISMA Flow Diagram (Page et al., 2021) for Review of Tests of Emotion Recognition in Children



1.4.2. Assessment of Emotion Recognition

The system of categorisation of tests of FER established by Paiva-Silva and colleagues (2016) is used to appraise the format and utility of tests of FER within the literature search. A critical analysis of the limitation of current tests will be provided and gaps in testing will be identified.

1.4.2.1. *Static tests using human face stimuli:* static tests using human face stimuli most often use a forced choice method whereby images of actors demonstrating an emotion must be labelled using only one of the six basic emotions (Beaudoin & Beauchamp, 2020). For example, Ekman and Friesen (1976) developed the Ekman-Friesen Pictures of Facial Affect test, the first widely used static test of FER using human face stimuli. This task requires participants to view images of actors demonstrating one of the six basic emotions and label the emotion. More recent static tests of FER have built upon a similar format to the Ekman-Friesen Pictures of Facial Affect test either by computerising the test (Langenecker et al., 2005; Nowicki & Duke, 2001; Tottenham et al., 2009); measuring the speed of responding (Langenecker et al., 2005; Rapport et al., 2002); alternating the intensity of the emotion shown (Herba et al., 2008; Nowicki & Duke, 2001); or using more ethnically diverse stimuli (Kessler et al., 2002). Some tests using this methodology have also incorporated a control task requiring participants to match or label stimuli such as animals or shapes (Herba et al., 2008; Peters et al., 2021) in order to control for the effect of differences in perceptual ability and/or processing speed.

An alternative method for static tests are matching tasks, where participants are required to match facial emotion stimuli to another image displaying the same emotional expression (Bowers et al., 1998; Herba et al., 2006). Matching tasks have the benefit of removing complex language requirements and proficiency in the testing language (Paiva-Silva et al., 2016).

Table 2 exhibits all the articles that fulfilled the inclusion criteria for this literature review and used static tests of FER with child participants. Most task used a “matching with emotion label format” whereby participants were shown a photograph of a facial expression of emotion and were required to match to one of the emotion label choices provided. Of the 12 tests included, seven were created for use on an

adult population group, but were used by researchers on child participants. Of the five FER tests which were specifically created for child populations only two were validated instruments; The Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 2001) and NEPSY-II (Korkman et al., 2012). The remaining three tests were designed for the study in question. The usage of unvalidated tests of FER in emotion recognition in children limits comparison between studies and may lead to bias within study findings (Paiva-Silva et al., 2016).

The NEPSY-II (Korkman et al., 2012) includes an affect recognition subtest which involves matching the target emotion to the same emotion in choice of four images of photographs of different children's facial expressions. While it has been found to be a promising measure of emotion recognition in children (Yao et al., 2018), this test is limited by floor and ceiling effects due to its design as a test aimed to capture performance both in very young and older children (Brooks et al., 2010). Floor effects emerge with the test being too challenging for very young children, whereas older children experience ceiling effects as the test is too easy for them.

DANVA (Nowicki & Duke, 2001) and DANVA2-CF (Nowicki & Duke, 2013) measure recognition of four of the six basic emotions ('anger', 'sadness', 'fear', and 'joy'). Researchers selected these emotions given that they were most frequently used in everyday interactions and according to previous research were more likely to familiar to children by age ten (Camras & Allison, 1985). DANVA uses adult face stimuli whereas the more recent DANVA2-CF consists of photos of children posing in high and low intensity of the four emotions being tested. With this tool participant must match the emotion in the photograph to the emotion label. DANVA has been criticised for having low correlations with other FER tests when assessing construct validity (Scherer & Scherer, 2011).

Table 2

Static Human Face Stimuli Tests of FER

Test Name	Population	Reference	Test Methodology
Ekman-Friesen Pictures of Facial affect test (Ekman & Friesen, 1976)	Adult	(Lawrence et al., 2015)	Matching with emotion label; adult facial emotion stimuli; six emotions; original FER test
Florida Affect Battery (Bowers et al., 1998)	Adults with brain injuries	(Tonks et al., 2007a)	Affect naming and matching with emotion label and image; adult facial emotion stimuli; six emotions
Diagnostic analysis of nonverbal accuracy (DANVA) (Nowicki & Duke, 2001) 22/05/2023 10:35:00	Child	(Nowicki et al., 2019) (Rapport et al., 2002)	Matching with emotion label; four emotions; adult facial stimuli
The Frankfurt Test and Training of Facial Affect Recognition (FEFA-2) (Bölte et al., 2006)	Adult	(Bölte et al., 2006) (Nowicki et al., 2019)	Matching to emotion label; adult facial stimuli; six emotions plus neutral face
Facially Expressed Emotion Labelling (FEEL) (Kessler et al., 2002)	Adult	(Khawar et al., 2014) (Lázaro et al., 2019)	Matching with emotion label; Japanese and White European adult facial stimuli; six emotions plus neutral
Facial Emotion Perception Test (FEPT) (Langenecker et al., 2005)	Adult	(Peters et al., 2021)	Matching to emotion label; speed of matching; adult facial stimuli; four emotions; animal categorisation as control for processing speed
Explicit Emotion Matching Task (Herba et al., 2006)	Child	(Herba et al., 2006)	Matching to emotion label and image; four tasks; adult facial stimuli; five emotions; four intensities of emotion; shape categorisation control
NEPSY-II Affect recognition (Korkman et al., 2012)	Child	(Korkman et al., 2012)	Matching with emotion image; child facial stimuli; six emotions
NimStim (Tottenham et al., 2009)	Adult	(Dede et al., 2021) (Leime et al., 2013)	Matching to emotion label; adult facial stimuli; four emotions
DANVA2-CF (Nowicki & Duke, 2013)	Child	(Nowicki et al., 2019)	DANVA2-CF Matching with emotion label; four emotions; child facial stimuli; two intensities of emotion
Unnamed (Franco et al., 2014)	Child	(Franco et al., 2014)	Matching with emotion label and image; White and East-Asian child facial emotion stimuli; five emotions
Unnamed (Crivelli et al., 2016)	Child	(Crivelli et al., 2016)	Matching with emotion image; cross cultural study; Mediterranean and Northern European adult facial stimuli; six emotions plus neutral
Reading the Mind in the Eyes of Children Test (Pahnke et al., 2020)	Adult	(Rueda et al., 2013)	Matching with emotion label; child facial stimuli eye region only; four emotions

1.4.2.2. *Dynamic tests using human face stimuli:* dynamic human face tests use video clip stimuli of actors demonstrating a emotional expression (Paiva-Silva et al., 2016). Three tests of FER using dynamic human face stimuli were identified as part of this literature search. All tests identified used dynamic stimuli in combination with static human face subtests. Tests followed a similar format whereby participants watched the video of an actor demonstrating the target emotion and were then requested to select the correct label for this emotion. Dynamic tests have the advantage of being more ecologically valid than static tests of FER (Paiva-Silva et al., 2016). Unlike with static tests of human face stimuli, all dynamic tests identified in this search were originally designed for children and young people. See Table 3 for information on the method used for emotion recognition tests.

Difficulties emerged with the validation of all three dynamic stimuli measures described in Table 3. Fridenson-Hayo et al., (2016), assessed convergent validity of their novel FER measure through correlation analysis with the FEFA-2 (Bölte et al., 2006). While this is a strength of their research, given that many studies using novel measures have not attempted to validate them, it should be noted that the FEFA-2 is a measure designed for adult participants and therefore is not the most appropriate measure to assess FER in children. Golan et al., (2015) assessed concurrent validity of the CAM-C against participant scores on the Mind in the Eyes task (Baron-Cohen, 1996). This task has faced criticism for its interpretation as a measure of multiple different socio cognitive constructs (Beaudoin & Beauchamp, 2020) possibly making it an unsuitable measure for assessment of concurrent validity. A strength of Taamallah et al.,'s (2021) paper assessing the validity of the Tunisian Test for Facial Emotion Recognition was their use of factorial analysis to validate their measure, having recognised the absence of a validated measure of FER for Tunisian children.

1.4.2.3. *Static tests using manipulated human face stimuli:* static manipulated tests utilise digital manipulation to either morph human faces to gradually increase or decrease the intensity of emotion displayed, or to alter the features of face stimuli (Paiva-Silva et al., 2016). Five tests are in this category; these are displayed on Table 4.

Table 3

Dynamic Human Face Stimuli Tests of FER

Test Name	Population	Reference	Test Format
Emotion Recognition Battery (Fridenson-Hayo et al., 2016)	Children with ASD	(Fridenson-Hayo et al., 2016)	Matching to emotion label; child facial stimuli; six basic emotions, six complex emotions
Cambridge Mindreading Face-Voice Battery for Children (CAM-C) (Golan et al., 2015)	Children with ASD	(Rodgers et al., 2021)	Matching to emotion label; age and ethnicity range facial stimuli; six basic emotions, six complex emotions
Tunisian Test for Facial Emotion Recognition (Taamallah et al., 2021)	Child	(Taamallah et al., 2021)	Matching to emotion label; age range facial stimuli; six basic emotions; varying intensity; photos and videos included

Table 4

Static Manipulated Human Face Stimuli Tests of FER

Test Name	Population group	Reference	Test Methodology
Unnamed test (Thomas et al., 2007)	Children	(Dede et al., 2021; Leime et al., 2013; Thomas et al., 2007)	Matching to emotion label; White adult facial stimuli; two emotions plus neutral; morphed progression
AFFECT (Gagliardi et al., 2003)	Children with William's Syndrome	(Gagliardi et al., 2003; Montirosso et al., 2010)	Matching to emotion label; adult facial stimuli; six emotions; morphed progression
FELT (Cecilione et al., 2017)	Child	(Cecilione et al., 2017)	Matching to emotion label; White adult facial stimuli; six emotions; morphed progression
The Fluid Emotions Test (Dyck et al., 2004)	Child	(Dyck, 2012)	Matching to emotion label; White European and Japanese adult facial stimuli; six emotions; morphed progression to another emotion; speed of identification
The Emotion Recognition Task (Kessels et al., 2014)	Children and adults	(Kessels et al., 2014)	Matching to emotion label; White adult facial stimuli; six emotions; Four levels of intensity

The majority of tests using manipulation of stimuli morphed the facial emotion stimuli from neutral to 100% intensity (Cecilione et al., 2017; Gagliardi et al., 2003; Thomas et al., 2007b). One test morphed stimuli from one emotion to another (Dyck, 2012) and one test provided several intensities of the same emotion (Kessels et al., 2014). Studies using this manipulated stimuli primarily investigate the intensity of emotion required in order for successful identification (Paiva-Silva et al., 2016).

Cecilione et al., (2017) created morphed stimuli that increased in expressivity in 10% increments from a neutral expression photograph to the emotion photograph. They found that from approximately 60% emotion intensity emotions become consistently more recognisable. Most tests identified in this category did not assess validity of their measures, however Cecilione et al., (2017) reported high test-retest reliability for the Facial Expression Labelling Task (FELT), and (Dyck, 2012) found adequate validity and reliability for the The Fluid Emotions Test.

1.4.2.4. *Dynamic tests using manipulated human face stimuli:* These tests of emotion recognition use manipulated stimuli, in video or moving images form, to determine ability in emotion recognition. Only one test is in this category by Herba and colleagues (2008) examined whether children had enhanced performance in identifying emotions from those familiar to them, for example teachers, as opposed to people unfamiliar to them. The researchers concluded that video stimuli facilitate children to identify emotions at a lower intensity than static stimuli.

1.4.2.5. *Static tests using computer generated stimuli:* FER tests in this category used computer generated images of emotions. One test is in this category. McKown and colleagues (2016) developed SELweb Emotion Recognition Test for children using digitised photos. In this test children are required to view the stimuli and apply to the correct emotion label. The stimuli used were digitised versions of children's faces.

1.4.2.6. *Static tests using drawings:* finally, researchers have used drawings or cartoon stimuli in their tests of emotion recognition (Howlin et al., 1999; Pons et al., 2004). These tests required children to point to the emotion described by the experimenter. Non-human stimuli have only been used in tests specifically designed

for children (Paiva-Silva et al., 2016), and have been criticised for their reduction in ecological validity. Despite this criticism, MacDonald and colleagues (1996) reported that children performed better on a test of FER using drawings than the Ekman-Friesen Pictures of Facial affect test (Ekman & Friesen, 1976) which uses human facial stimuli. This suggests that drawings and cartoons may have the benefit of increased familiarity for children (MacDonald et al., 1996). In a review Paiva-Silva et al., (2016) recommended more research of the efficacy of tests using drawing or cartoons, with consideration of the relationship of performance in these tests and real word social functioning.

1.4.2.7. *Summary and critique of tests of FER in children:* most tests of FER identified in this literature review have been developed as variations of the adult test, the Ekman-Friesen Pictures of Facial affect test (Ekman & Friesen, 1976). Despite the development of dynamic, and static manipulated tests specifically for children, most studies identified in this review used static test stimuli, often using images from Ekman and Friesen, (1976). Most tests used a similar format; a directive forced choice matching of emotion word label or emotion image. In general, validated tests included in the literature review were found to have good reliability and validity (Paiva-Silva et al., 2016), however many tests included were not formally validated. The use of forced choice methodology raises the question of ecological validity; is a child's FER ability during a testing session when their attention is explicitly drawn to facial emotions, directly comparable to how they function in a real-world environment where there are competing salient stimuli to consider.

1.4.3. Culturally Fair Test of FER

The majority of tests of FER identified in the literature search did not reference cultural considerations during test development nor refer to the ethnicity of facial stimuli used. One exception was The Tunisian Test for Facial Emotion Recognition (Taamallah et al., 2021) which was developed specifically for use on Tunisian children. Despite this, the authors did not identify steps taken to ensure the cultural fairness of their test for this group. However, the use of Tunisian ethnicity actors to develop the test materials indicated that novel facial stimuli was developed for the test. One test the Cambridge Mindreading Face-Voice Batter for Children (CAM-C;

Golan et al., 2015) specified the use of faces of people from a range of ethnicities as the FER stimuli. Kessler and colleagues' (2002) Facially Expressed Emotion Labelling Test specified the use of both White European and Japanese facial stimuli. Crivelli and colleagues (2016) described using FER stimuli from actors with Northern European and Mediterranean ethnicities, Franco et al., (2014) used a combination of White and East-Asian facial stimuli.

Ekman's pioneering research in the 1960s established that the six basic emotions endure cross-culturally (Ekman, 2003). This remains widely accepted in the literature and these emotions continue to be included in the majority of tests of FER. Despite the almost singular focus on these emotions in tests of FER there is little evidence in the literature of the exploration of other factors which might enhance the cultural fairness of tests of FER. This is contrary to indications that even when there is evidence for the universality of a cognitive construct the means in which it is expressed by people of diverse cultures can vary (Fernández Abe 2018).

Matching tasks requiring the participant to identify faces with similar emotional expression is a common format in existing tests of FER. This method has the advantage of enhancing the cultural fairness of the test by reducing the reliance on language. Language intensive tests are a possible limitation for participants who are not fully proficient in the language of testing (Fernández & Abe, 2018).

In order to improve the cultural fairness of cognitive tests, Fernández and Abe (2018) suggest that developing new culturally fair and psychometrically robust tests is preferred over the adaptation of existing tests in order to make them more culturally appropriate. They highlight that many existing cognitive test are developed not for clinical but for research purposes and as a result can be unsuitable for use in clinical settings. The development of novel test stimuli also allows for culturally fair stimuli and a move away from the limitations inherent in using images of real world faces.

1.4.4. Principles of Good Tests of Social Cognition in Children

The importance of assessment of social cognition has been highlighted in recent years given its prevalence in neurodevelopmental, acquired, and environmental

disorders (Beaudoin & Beauchamp, 2020). As a result, assessment of social cognition has now been included as a core cognitive area in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013). Despite this there has been little research addressing the development of tests of emotion recognition for children. Current tests of emotion recognition in children have either been developed for adults or rely upon the same methods as tests for adults but instead use child facial emotion imagery. The development of new tests of emotion recognition for children is a new area of research. In order to guide development of new tests Beaudoin and Beauchamp (2020) described principles for good tests of social cognition, outlined below.

1.4.4.1. *Clarity of sociocognitive process being assessed:* a limitation of currently available assessments of social cognition is the lack of consensus definitions and uniform taxonomy. This has resulted in confusion of which sociocognitive construct a test is measuring (Beaudoin & Beauchamp, 2020). For example, research has a history of equivalating poor performance on a test of FER to a deficit in empathy or theory of mind, which in themselves were not measured in the test. Therefor clarity in the description and aim of the sociocognitive test is important.

1.4.4.2. *Assessment of social cognition should be developmentally appropriate:* tests of social cognition for children must have a basis in developmental norms. Consideration should be given to both the norms for the acquisition of the targeted sociocognitive function but also the acquisition of other cognitive functions which may be required in the assessment task. Many tests of emotion recognition rely upon verbal and written comprehension and response. Such tests should only be used at an age when these skills have been developmentally acquired, resulting in floor effects (Schworer et al., 2021). Similarly, consideration of other cognitive processes such as perceptual functions and working memory should be given, in particular if tasks are aimed for very young participants.

1.4.4.3. *Sociocognitive measures should have good psychometric properties:* due to the relative recency of sociocognitive testing, gold standards of good sociocognitive tests are not yet available (Beaudoin & Beauchamp, 2020). Despite this

consideration should be given to reviews of available assessment measures in order to establish good psychometric properties.

1.5. Review of Literature – Game-like Tests of Cognition

Research on the development of game-like tests of other cognitive functions will be explored in this section to evaluate the utility of creating a game-like version of a FER task. This literature review revealed that much of the research published in this area is focused on computerised game-like tests. With this in mind the following subsection will focus on the rational for selection and development of game-like tests and the success to which these tests appear to measure the cognitive construct in question.

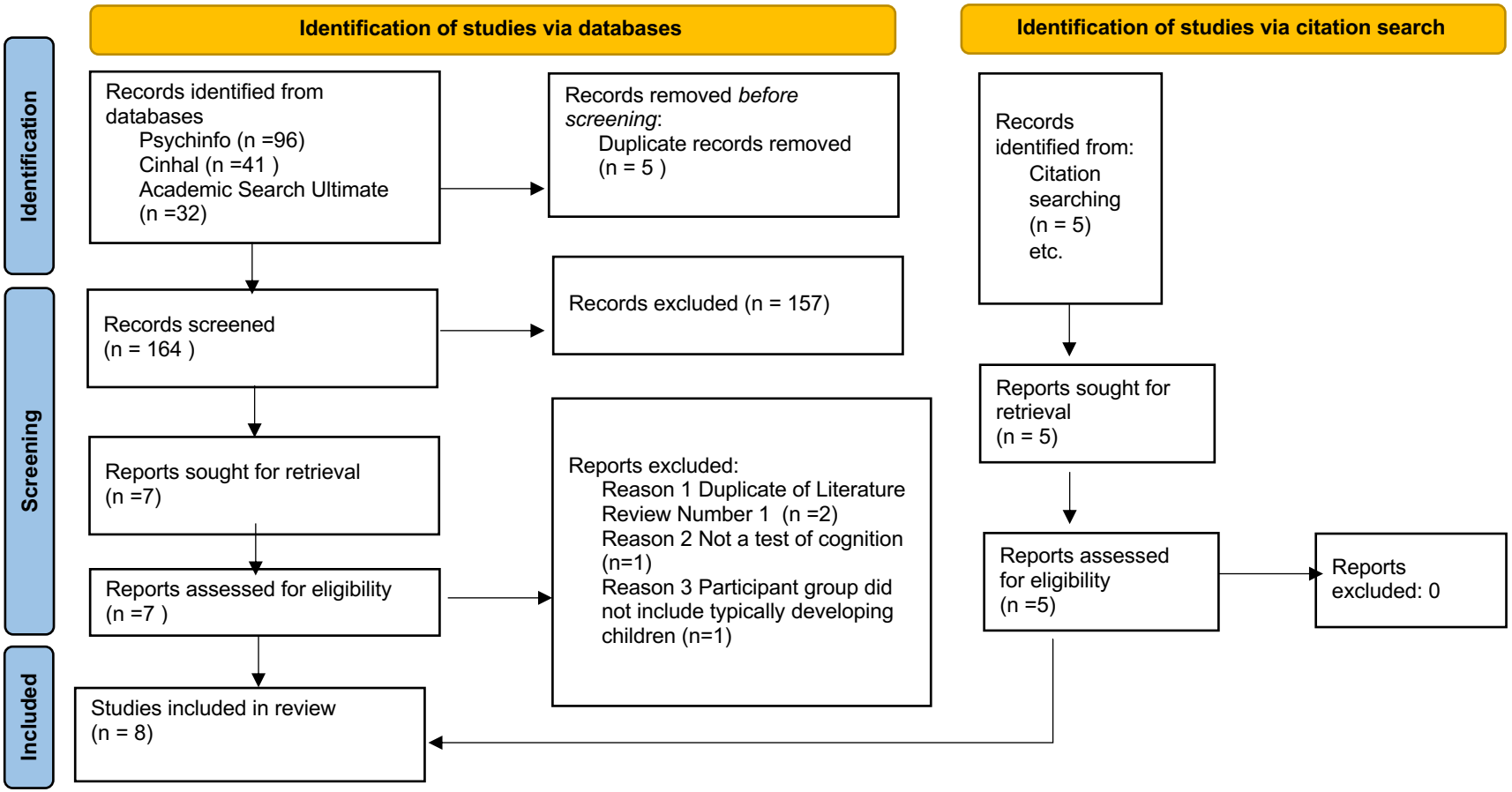
1.5.1. Search Results

The results from literature search B investigating the literature base for “game-like cognitive tests for children” are displayed in Figure 2 below. The total number of articles identified through search of databases were 169. Duplicate papers were removed, and titles and abstracts were manually screened for eligibility and relevance to research questions. Full text of articles were read for papers which appeared to meet inclusion criteria. Citation searching of these papers identified an additional five papers.

Eight papers were determined to be eligible for inclusion. The following section provides an overview of the evidence from these studies.

Figure 2

PRISMA flow diagram (Page et al., 2021) for Review of Game-like Cognitive Test for Children



1.5.2. Definition of Games

Various philosophical positions have produced distinctive frameworks attempting to define and understand the use of games. Despite the breath of philosophical opinion much of this work focuses on computer games and sport (Nguyen, 2017). Following a review of the literature on game-like tests of cognition, it seems that the 'magic circle' framework of games (Salen & Zimmerman, 2003) is the most appropriate lens to view game-like test of cognition through. This approach conceptualises game-play as a prop which can be used in order to interpret or understand real life (Nguyen, 2017). Game-like tests of cognition also likely benefit from the novelty and increased engagement of game-like tests in enhancing motivation of children to engage with testing (Hawkins et al., 2013; Staphorst et al., 2017)

1.5.3. Development of Game-like Tests

Within studies eligible for the literature review there emerged two main approaches for the development of game-like tests. The first was a responsive approach in which existing simple individual computer games were analysed by a team of researchers and classified, through a standardised process, into the cognitive categories explored by the task. This approach was used by Martinovic et al., (2015) who compared 15 computer games with performance on subtests of the NEPSY-II in order to investigate the utility of computer games as potential cognitive tests. Computer games of interest were selected on the basis of utilising cognitive skills contained in NEPSY-II subtests. Martinovic and colleagues (2015) study was comprehensive however administering both the NEPSY-II and the novel computer games resulted in study sessions of up to four and half hours, leading to fatigue of participants and possibly impacting performance on the measures.

The second approach was more frequently used: researchers developed novel videogames, specifically designed to target cognitive functions. Song et al., (2020) used this approach in the development of their mobile computer game application 'ConCon' to assess cognitive control in children and young adolescents aged between 9 and 16 years. They developed the video games to correspond to tests of attention, working memory, cognitive control, and cognitive execution. Mukherjee et al., (2020) followed a

similar approach in their development of DEEP (Developmental assessment on an E-Platform) a game-like tool for cognitive assessment. They used a team of experts from psychological, medical and neuroscience backgrounds to develop the games included in the test. Kerns (2000) also used this technique; similar to Mukherjee et al., (2020) they took inspiration from games already enjoyed by children and manipulated this to ensure that it targeted prospective memory. By doing this they were able to target the desired cognitive construct and ensure the game was enjoyable enough to provide motivation for the child.

1.5.4. Evaluation of Game-like Tests

Research evaluating game-like tests of cognition primarily focused on comparison to existing measures of cognition and the acceptability of the test. The subsection below explores the evaluation of game-like tests using these criteria in further detail.

1.5.4.1. *Concurrent Validity:* the most common method of evaluating game-like tests was to compare performance on these tasks to existing standardised measures of the same cognitive construct. Martinovic et al., (2015) carried out correlation analysis between computer game results and the NEPSY-II subtest that researchers thought evaluated similar cognitive skills. Eleven of the 15 games had high correlations with the NEPSY-II subtests indicating that computer games might be used to assess cognitive skills. As a result, the researchers concluded that game-like cognitive tests are potentially valid and reliable. The task 'ConCon' was compared to traditional neuropsychological tests including the ten core subtests of the Korean WISC-IV (Kwan et al., 2011), the Stroop task (Stroop, 1935), and a self-report questionnaire of executive functioning (Song et al., 2020). Researchers identified a correlation between the game and standardised measures; and determined that the game-like 'ConCon' task was a valid and reliable measure.

Researchers investigating the efficacy of DEEP (Mukherjee et al., 2020) found that non-specialists could deliver the test, and that scores correlated with the Bayley's scale of infant and toddler development. Gómez-Tello et al., (2022) assessed and compared the

scores produced by the TOWI (Rosetti et al., 2017) a game-like neuropsychological screening task, and standardised cognitive tests. They found that TOWI was a sensitive measure of cognition and that choices in design while gamifying the test did not impact on their ability to measure cognitive performance.

1.5.4.2. *Acceptability of and engagement with the task:* the DEEP game-like test (Mukherjee et al., 2020) was piloted prior to further development in order to ensure acceptability of the tests for both children and their careers, deliverability of the test and whether it could discriminate cognitive ability (Bhavnani et al., 2019). It is interesting that despite increased engagement in game-like tasks being a major factor in the development of the tests no other study asked participants directly whether they were engaging or enjoyable. Research has historically relied upon observational data to establish that game-like tasks are engaging for young participants (Mukherjee et al., 2020; Song et al., 2020).

1.5.5. Critical Analysis of Game-like Tests

1.5.5.1. *Limitations of computerised gamification of cognitive tests:* the literature indicates that researchers should be cautious as to whether the game-like tasks are assessing the targeted cognitive function. Martinovic and colleagues (2015) found correlations between computer games and NEPSY-II subtests ‘Clocks’, ‘Design Fluency’ and ‘Inhibition’, which they had not anticipated to be related. One possible explanation for this is that games, in general, have a level of complexity which draws upon a very broad range of cognitive functions and prevents adequate manipulation over certain conditions (Granic et al., 2014). Similarly Song et al., (2020) noted that the multidimensionality of computerised and game-like assessment should be monitored to ensure the validity of the test.

1.5.5.2. *Strengths of gamification of cognitive tests:* game-like tests are intended to promote participant engagement and motivation in testing (Lumsden et al., 2016; Song et al., 2020). In addition, they may improve the ecological validity of the task (Lumsden et al., 2016; Pitchford & Outhwaite, 2016). In realistic circumstances, a game-like task

allows the participant flexibility in their responses and to experience the effects of the choices they make within the task. Studies examining the impact of gamification have found that both game-like and non-game tests can provide comparable results (McPherson & Burns, 2007, 2008) with the game-like version being more enjoyable according to subjective reports (Hawkins et al., 2013).

1.5.6. Recommendations for Development of Tests

The existing literature on game-like cognitive tests has indicated a number of recommendations for future development of game-like tasks. These are outlined in the subsection below.

1.5.6.1*. Identify what reward of the game-like test is:* improved participant engagement in game-like tasks is likely a result of enhanced motivation by encouraging participants to complete the task (Lewis et al., 2016). Therefore it is important for test developers to identify what the reward of the task is; for example winning, kudos, feedback etc. Increased engagement is particularly important for child participant groups as children are more likely to experience boredom during testing (Staphorst et al., 2017).

1.5.6.2*. Use existing games which are known to be enjoyable for children:* game-like tasks increase motivation of participants completing tasks that may otherwise be considered demotivating due to their repetitive or frustrating nature. Previous researchers who have created game-like tests have adapted existing games like “spot the difference” (Martinovic et al., 2015), or a driving game (Kerns, 2000) around the principles of an existing test of cognition to enhance both motivation and enjoyment of the task.

1.5.6.3. *Match tasks to the developmental level of child*: game-like tasks can increase accessibility and usability of the task for its intended audience. Development of new tasks should match the level of complexity to the target population. Game-like tests for younger participants have the added benefit of preventing boredom and improving conditions of testing for young children (Tenorio Delgado et al., 2016).

1.6. Summary

As evidenced from this review, current tests of FER used with children, were developed from adult tests of FER, with few or no adaptations for child usage. This is despite the gradual development throughout childhood and adolescence of social cognitive skills required for FER (Beaudoin & Beauchamp, 2020). Children's ability to identify and label emotions develops gradually throughout childhood (Kolb et al., 1992), with further refinement of these skills continuing into adulthood (Khawar et al., 2014). Given the range of social, emotional, and behavioural difficulties that can arise from deficits in FER, it is vital that tests of FER accurate capture difficulties, in order to identify young people with such difficulties. More sensitive tests of FER for children might function to capture deficits in FER with greater precision.

Possible differences in how children and adults engage with tasks is another reason why it is important to develop tests of FER specifically for children. For example, adult engagement in tests is influenced by the face validity (the test appears valid to the examinee) (Nevo, 1985). Tests of FER undoubtedly have high face validity given the format used. For children, however, other factors may be more important for engagement. Child engagement in tests can be enhanced by a game-like format (Lumsden et al., 2016; Song et al., 2020). The literature suggests that there are currently no game-like tests of FER for children.

Review of the literature on tests of FER outlined above has also revealed that existing tests of FER have largely failed to take into account the cultural considerations in both methodology and stimuli used. Development of a novel test of FER would allow for cultural considerations to be incorporated in multiple aspects of the testing, something that is made difficult when adapting existing tests (Fernández & Abe, 2018).

1.7. Present Study

Given the limitations of current tests of FER outlined above, this study aimed to develop a test of FER which accurately captures a child's ability. To encourage engagement a game-like procedure was included in the testing format of the task. Novel culturally-neutral stimuli were used to ensure all children could relate equally to the task regardless of cultural background. The novel FER stimuli was also administered in the established testing format in order to facilitate comparison to traditional tests of FER.

1.7.1. Study Aims and Rationale

The current study is an exploratory, first phase development of a new test of emotion recognition for children; *the Alien Quiz*. The aim of this study is to develop a new, engaging, enjoyable, and culturally fair test of FER for children of primary school age. This novel test was created with the objective of being developmentally appropriate for children aged 7-11. It is hoped that the use of a game-like procedure will function to help maintain the attention and reduce the stress of testing for children. Reducing the negative impact of testing is of particular importance if a test is to be useful in clinical practice with children who have developmental delay.

As this is the first phase of development of this test, the study was conducted with children attending mainstream primary school. The aim of this was to determine whether typical response could be established from this group's approach to the test materials, in addition to identifying whether the test was appropriate for those in the target age range. This phase of development also aimed to determine whether children found the task engaging, and whether there was a relationship between children's responses and existing measures of interpersonal competence.

1.7.2. A Game-like Procedure

The current study builds upon Pavitt's (2017) "Alien Game", a game-like test of conception formation for children. Pavitt's (2017) study was influenced by a test of category processing developed by Alderson-Day & McGonigle-Chalmers (2011). Similar

to Pavitt (2017) and Alderson-Day and McGonigle-Chalmers (2011)'s studies, the current test utilises a 20-Questions style procedure in order to gamify the FER task. During the Guess the Alien subtest, participants are required to ask yes/no questions to identify the target character of the opponent player and win the game. Twenty-four novel "Aliens" were designed in order to be culturally neutral, represent different emotional expressions, and be visually appealing to children. These were provided on laminated cards which the children could flip over. Three other subtests, using traditional FER test formats, were also created with this novel stimulus.

1.8. Research Questions

This study aims to answer the following research questions:

- Do typically developing children detect and draw upon emotional expression as a means of categorisation in the Guess the Alien subtest?
- Do typically developing children apply similar strategies during testing with the Alien Quiz?
- Do primary school aged children engage well with the Alien Quiz as a test of FER?
- Is there a relationship between performance on the Alien Quiz and a test of real-world social functioning?

2. EPISTEMOLOGY AND METHODOLOGY

2.1. Section Overview

Epistemology is concerned with the theory of knowledge; specifically, epistemology explores how awareness of knowledge is gained and the extent of this knowledge (Willig, 2008). Epistemology influences the theoretical perspective of research, which in turn determines the methodology which informs the methods (Crotty, 1998). As a result, a clear understanding of the epistemological foundation of research and the methodological requirements is necessary to assess whether the research objectives have been met (Willig, 2008). A critical realist position was adopted within the present research. The section below explores the different epistemological positions, the reasons why a critical realist approach was chosen and the study methodology.

2.2. Epistemological Approaches

Crotty (1998) identified three main epistemological approaches: constructionism, objectivism, and subjectivism. Constructionism relates to meaning that is created through human engagement, and posits that there is no single truth and no meaning without a human lens (Al-Ababneh, 2020). Subjectivism refers to meaning that emerges from anything other than the entity with which it was attributed (Crotty, 1998). A subjectivist approach holds in mind that knowledge gathered from observations may be fallible and that observations are dependent upon the theory held by the observer. Finally, objectivism suggests that meaning exists separately from human consciousness (Crotty, 1998), which in turn implies that social phenomena exist outside the social actors (Saunders et al., 2009). In taking this stance it is considered that there are no major barriers to collecting truthful knowledge about the world.

Emerging from the epistemological approach are theoretical perspectives which are philosophical stances which advise methodology (Crotty, 1998). Saunders and

colleagues (2009) identified four main theoretical perspectives: realism, interpretivism, positivism and pragmatism.

2.2.1. A Critical Realist Approach

The current study utilised a critical realist philosophical position. A critical realist approach conceives that an objective reality exists independently of human perception, however a person's own subjective account of the world is obtained from the historical, social, and political environment in which they exist (Bhaskar, 2010). In such an approach the researcher must maintain an awareness that individual interpretations of the phenomena are open to error and thus a critical stance is required (Al-Ababneh, 2020).

Critical realism stems from a subjectivist epistemology stance and acknowledges that an independent external world is in existence, but also allows for the understanding that knowledge is altered through social construction including the wider social, political and historical context (Bhaskar, 2010). In the context of this thesis, an attempt was made to explore and assess a specific phenomenon, facial emotion recognition, within a specific time, space, and social context that the research considers to be separate and independent of personal experience. A critical realist approach allowed for the use of a theory-dependent method which it is anticipated will lead to findings which will contribute usefully to the evidence base for assessment of emotion recognition in children.

In the analysis and interpretation of the data, the researcher maintained an awareness that the observations and findings made have the potential to be prone to bias and error (Trochim, 2001). In keeping with this subjectivist approach of fallibility, knowledge was considered not as fact but as something to be interpreted with caution and awareness of its limits was maintained.

2.3. Design

The current study employed a quantitative mixed methods approach, using a within-participant and cross-sectional design. An exploratory approach allowed for participant strategies and performance characteristics on the Alien Quiz to be evaluated, while also appraising and developing methods of measuring responses to the task. Participant responses in the Alien Quiz were measure by comparing frequency and quality of questions related to emotion recognition and quantifying this data into a scoring criterion. A cross-sectional design was used to compare scores the Alien Quiz against established teacher-rated social cognitive scores to determine ecological validity. A within participant design allowed for comparison of scores within different subscales of the Alien Quiz in order to assess internal reliability.

2.4. Inclusion and Exclusion Criteria

The inclusion criteria for this study were children, aged between 6-11 years old, who understand verbal and written English, and who attend a mainstream primary school. This age range was selected as early to mid-childhood is the developmental period in which facial emotion recognition is established (Beaudoin & Beauchamp, 2020). No exclusion criteria were used; any child who met the inclusion criteria was eligible to participate in the study.

This study aimed to recruit 20-30 participants; this calculation was made on the basis of participant numbers used to develop similar sociocognitive tasks with children (Golan et al., 2015; Pavitt, 2017).

2.5. Test Development

The Alien Quiz was developed in order to assess the children's ability recognise different facial emotions. This test comprises four subtests. Subtests were developed to represent a continuum of FER assessment ranging from an indirect (self-generated)

measure of FER, towards more directive (cued and perceptual matching) identification of FER. The first subscale Guess the Alien aims to identify self-generated FER ability through a novel game-like approach to testing FER in children. The following three subscales adapt existing formats of tests of FER using child friendly stimuli; in these subtests self-generated FER ability through confrontation, cued FER ability, and perceptual matching ability of FER are assessed. Four short trials of Guess the Alien were used, alongside one trial each of the three other Alien Quiz subscales. This aimed to provide that the test was developmentally appropriate for participants at the lower end of the age range who have lower sustained attention ability (Betts et al., 2006). Language used in testing and included in the written instructions was accessible for children ages 6-11 in keeping with developmental appropriateness of the test.

The stimuli for the test consisted of 24 numbered Aliens arranged on individual laminated cards that could be flipped over. Each alien had one of seven emotional expressions in addition to non-emotion related visual characteristics such as a legs and different colours. The appearance of selected characteristics was standardised to the same frequency in order to allow for comparison of questions related to emotion and those related to non-emotion traits (see Appendix A for stimuli development scheme with breakdown of assigned characteristics). Selected characteristics were chosen to be accessible for children to name in keeping with their developmental stage.

2.5.1. Development of Facial Emotion Stimuli

The six basic emotions, outlined by Ekman and Friesen, (2003) and widely utilised across tests of FER (Paiva-Silva et al., 2016), were selected for use in this test alongside a neutral expression. In order to create stimuli fitting with a game-like test facial emotions were adapted to a cartoon form on alien characters. There is evidence that cartoon style drawings may be more developmentally appropriate for children and are more familiar to them (MacDonald et al., 1996). The use of aliens also allowed for the creation of more culturally neutral stimuli, which serves to control for the suggested in-group advantage for the recognition of emotions across culture (Elfenbein & Ambady, 2002). Aliens were custom drawn by the researcher using Adobe Illustrator for desktop

(version 26.3.1.; Adobe, 2022); Adobe Stock (Adobe, 2022b) emotions were selected and used as a template for the emotion stimuli and were adapted to fit emotion criteria. This approach ensured that the test presented to the children was novel, ensuring no practice effects were present. It also allowed characteristics to be manipulated and altered to create facial expressions consistently recognisable to the six basic emotions.

Stimuli from Ekman's photographs of facial emotions (Ekman & Friesen, 2003) were analysed by the researcher and the research supervisor in order to inform key characteristics to include in the development of the emotions on the aliens. Key characteristics were compiled for each emotion and cross referenced with the existing literature. These traits informed the creation of the emotion stimuli; see Figure 3 for emotion stimuli and key traits.

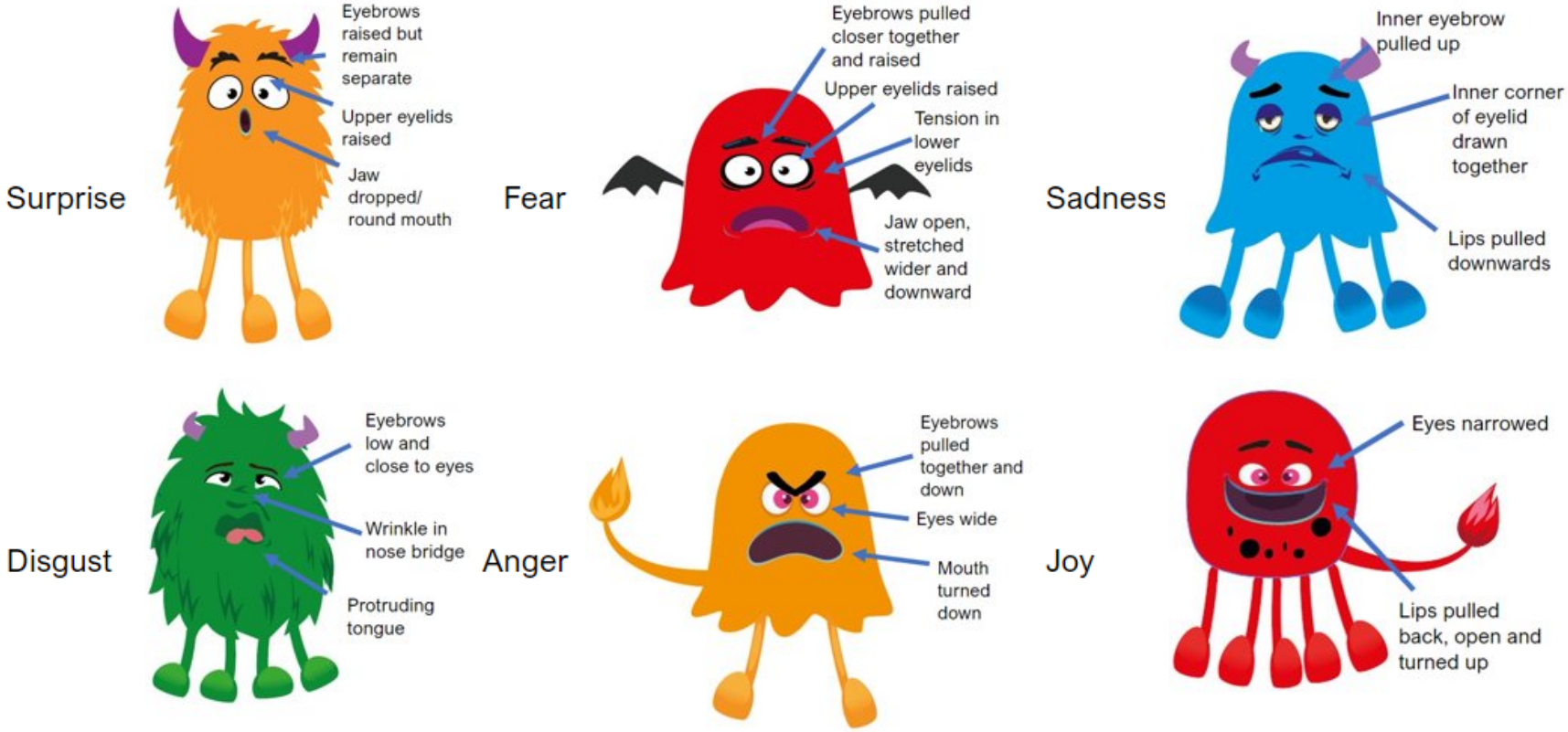
A three-phase piloting strategy was used to create the final version of the facial emotions included in the study. In the first phase the researcher and the research supervisor examined the first prototype emotion stimuli and assessed the following criteria:

- whether the target emotion was recognisable
- whether the traits of the emotion were visible
- whether other features of the emotion stimuli (the Alien) altered the interpretation of the emotion in the alien.

The researcher then made the suggested change to the image; see Appendix B for example of alterations made during the pilot phase. Following this, the emotion stimuli were piloted with three adults known to the researcher to establish whether the emotions were recognisable. Finally, the emotion images were piloted with two children known to the researcher, and in the age range of the participant group, to establish whether the emotions were recognisable to children.

Figure 3

Example Emotion Stimuli and Key Emotion Traits



Feedback was sought and received from both the adult and children who viewed the material. This feedback was discussed with the research supervisor and changes were made in line with the literature base for traits of facial emotions. Changes made during the piloting phase included: removal of movement lines in wings as this could be interpreted as excitement; removal of teeth as variable as this altered emotion expression; eyebrows raised for ‘fear’ emotion; wrinkle in nose bridge for ‘disgust’; general changes to ensure consistency of all characteristics.

During piloting, difficulties emerged with creating a cartoon version of ‘disgust’. In adults, a protruding lower lip is a key feature of facial expression of ‘disgust’ (Ekman & Friesen, 2003). Prototypes of the novel stimuli failed to accurately convey this feature in the cartoon images. An alternative, a protruding tongue, was considered appropriate in discussion with the research supervisor. This facial emotion trait relates more closely to a child’s experience of ‘disgust’ and so was considered an acceptable alternative (Missaghi-Lakshman & Whissell, 1991). This emotion image was correctly identified in the child pilot.

2.5.2. Development of a Game-like Strategy

In keeping with findings that gamelike tasks improve both motivation and engagement of the task (Lumsden et al., 2016; Song et al., 2020) and also the ecological validity of the task (Lumsden et al., 2016; Pitchford & Outhwaite, 2016) Guess the Alien was developed with the goal of using a game-like approach to a sociocognitive measure. A structure similar to the established children’s game “Guess Who” and “20 Questions” was adapted for Guess the Alien’’ the first subscale of the FER task. A similar tool “the Alien Game” has been investigated as a measure of concept formation (Alderson-Day & McGonigle-Chalmers, 2011; Pavitt, 2017), but to my knowledge the proposed research is the first investigation of such a measure as an assessment of emotion recognition.

Similar to the approach used by other game-like tasks of cognition (Kerns, 2000; Martinovic et al., 2015), the strategy used in this study was to adapt an existing game already enjoyed by children. The adaptation of an existing game also aimed to

function to enhance the motivation for the task. Children were told that to “win” the game they are required to complete the task by asking as few questions as possible, and that 10 was the most questions that they could ask. General verbal feedback was provided to all participants to enhance motivation in playing the game and increase reward of the game.

All 24 cards with alien characters were placed in front of the participant. The researcher held a card and indicated that the child had to guess the alien that appeared on this card by asking questions about the alien images. Participants were informed of the rules of the game; they could only ask yes, or no questions and were encouraged only to ask questions about the numbers on the cards when identifying the target alien. Participants were instructed to flip over cards of aliens that, as indicated by the response to their question, were not the target alien. Four trials of the task were completed.

2.5.3. Adaptation of Traditional Format FER Tests

The remaining three subscales created in the Alien Quiz FER test battery were Emotion Naming (free), Emotion Naming (cued) and Emotion Matching.

2.5.3.1. *Emotion Naming (free):* in this task, participants were shown, one at a time, an Alien card containing one of six basic emotions. Participants were requested to name how the alien was feeling. Participant response was recorded verbatim on the response sheet. One point was given per each emotion correctly identified. Close synonyms were awarded full points; the research supervisor and Collins Thesaurus (Collins, n.d.) were used to confirm whether a synonym emotion was adequate. Scores on this subscale could vary from 0-6 points.

Although less frequently utilised as a test of FER, affect naming likely provides increased ecological validity as a test of FER as participants are not selecting from pre-established label options. This test methodology has previously been used in the Florida Affect Battery (Bowers et al., 1998), whereby participants are required to name the facial expression on faces displaying ‘angry’, ‘joy’, ‘sad’, ‘frightened’, and

'neutral' faces. This test was originally validated on adults (Bowers et al., 1998) but has been used on child populations aged 9-15 (Tonks et al., 2007a).

2.5.3.2. *Emotion Naming (cued):* participants were again shown, one at a time, an Alien card containing one of six basic emotions. Participants were invited to choose the correct emotion from a selection of four possible options. For example, participants were asked "is the alien feeling angry, afraid, happy, or surprised?" Each emotion appeared four times in the response options. Responses were recorded verbatim, and one point was given per correct response. Scores on this subscale were from 0-6 points.

The majority of tests of FER for both adults and children use this method (see section 1.4.2. for greater detail). In these tests participants are required to select the correct emotion label for the facial expression of emotion shown. Of the measures that employ this methodology which have been validated there is evidence that this is a valid and reliable test format for both adults (Tottenham et al., 2009) and children (Nowicki & Duke, 2001)

2.5.3.3. *Emotion Matching:* participants were instructed to select the Alien cards which matched the same emotion or were "feeling the same way" as the target alien. All 24 emotion stimuli were placed in front of the participant. The participant was instructed to collect all aliens that matched the emotion of the target alien. The number of the stimuli identified by the participants were recorded. One point was allocated per correct emotion stimuli/alien identified. Scores on this subscale were from 0-15 points.

Existing tests of FER frequently use affect matching methodology in conjunction with another subtest of affect recognition (see 1.4.2. above). One existing test of FER was identified which used only an affect matching methodology the NEPSY-II Affect recognition (Korkman et al., 2007). This test was found to have adequate technical qualities and item difficulty in young children (Yao et al., 2018).

2.5.3.4. *Total FER:* a total FER score was developed which combined participant scores on the three adapted tests of FER described above.

2.6. Test Battery

The following items were used during the test procedure in this study:

- Child Information Sheet (Appendix C)
- Child Consent Form (Appendix D)
- Child Debrief Sheet (Appendix E)
- Record Form (Appendix F)
- Instruction Sheet (Appendix G)
- Peer Problem and Prosocial Subscale of the Strengths and Difficulties Questionnaire (SDQ, Appendix H)
- The Alien Quiz (Appendix I)

2.6.1. Assessment of Current Social Behaviour

Two subscales of the SDQ (Goodman et al., 1998), were completed by the year group teacher of the participant. The Peer Problems subscale was used to assess participant difficulties in their relationships and interactions with peers; the prosocial subscale was delivered to evaluate the prosocial resources, such as social skills and competencies that the participant possess (Silva et al., 2015). Although not a measure of social cognition, SDQ subscale scores provide an insight into the participants real world social behaviour. The SDQ has robust psychometric validity and reliability and is used widely in research (Stone et al., 2010).

2.6.2. Test of Emotion Recognition

The novel FER test battery 'The Alien Quiz' was administered in order to measure FER ability. This included four subscales; Guess the Alien, Emotion Naming (free), Emotion Naming (cued) and Emotion Matching.

2.6.3. Acceptability of Task

visual analogue scale will be used to assess whether the participant enjoyed the task.

2.7. Ethics

Ethics approval for this study was received from University of East London's School of Psychology Ethics Committee (Appendix J).

This study was conducted in a London primary school. Head teachers and special education needs coordinators (SENCO) of multiple London based schools were approached, via email and research flier (Appendix K) for participation in this study and another similar study. One SENCO responded that their school was interest in the research occurring in their primary school. The head teacher was provided with a copy of the study information sheet for schools (Appendix L) and following review of this agreed to participate in the study. The head teacher was asked whether they would prefer for parents to opt their children into the study, or if they would prefer parents to opt out their children from the study. The head teacher selected the opt out approach and completed an *In Loco Parentis* consent form (Appendix M) for the children who were not opted out by their parents or guardians. Two months prior to data collection the school included the research study in their school newsletters informing caregivers that the study would be taking place. One week prior to data collection the school sent home information sheets and opt out consent forms (see Appendix N) to Year 3 and Year 5 pupils. On the first day of data collection the SENCO provided a list of children who were eligible to participate and had not been opted out of the study; this ensured that no child who was opted out of the study by their guardian was approached to take part in the study.

At the beginning of data collection for each year group the researcher told the students about the research and outlined that students would be contacted throughout the day and invited to take part in the research. Students were offered the opportunity to ask questions about the research in a group setting, prior to meeting with the researcher individually. The researcher met with each eligible child individually and read and discussed the information sheet with them. Children were given the opportunity to ask questions about the research and their participation and were reassured that participation was optional and declining to participate would have no negative consequences. Children who agreed to participate were requested to sign the child consent form. Complete information about the study was provided to

both child participants, their parents, and the head teacher at the participating school. No deception was used in this study. Participants were provided with debrief sheets following completion of the study.

Participants and their guardian were informed of their right to withdraw their participation in the study at any time during testing by informing the researcher. They were also informed of their right to withdraw their data from their study up until two weeks after testing occurred, by emailing the researcher's university email address. If a child appeared distressed or fatigued during testing, they were offered a break or the option to end their participation. No children appeared distressed during testing. No children or guardians requested the withdrawal of their data.

2.7.1. Confidentiality

Consent forms which contained the participant's name and signature were stored in a separate folder on the researcher's password protected university OneDrive account, no other users had access to this folder. Test records were stored separately from consent forms. Each participant was allocated a unique code which was written on the bottom of their consent forms, and testing record, and the teacher rated SDQ. SDQ's contained the child's name of the top of the page to ensure that the teacher's response corresponded to the correct participant. Following return of the form the teacher the child's name was removed to maintain confidentiality. Consent forms were retained until analysis of data to allow for data to be withdrawn at the request of the participant or their guardian. Participant code, age, sex, and test data were included in excel and SPSS databases for analysis. These databases were stored on the researcher's password protected university OneDrive account.

2.8. Procedure

The test procedure was administered in a quiet room in the school. The participant was seated facing the researcher on the opposite side of the desk. The Alien Quiz was laid out in front of the participant on individual laminated cards numbered from 1-24; see Appendix I for image depicting the layout of the game. Children were asked whether they had played the game Guess Who before, they were then read a

set of standardised instructions on the game (see Appendix I). In the instructions the participants were instructed to ask “Yes” or “No” questions about the Aliens to find the target Alien in the fewest possible questions. No questions on the numbers attached to the Aliens were permitted, unless it was to inquire about the target alien. Children were instructed to flip over the Aliens which were eliminated after each question. The researcher did not correct participants if they flipped over Aliens incorrectly, but noted with pen and paper if errors were made relating to FER. Participants questions, comments, and behavioural observations were recorded using pen and paper. Testing aimed to complete four trials of the Guess Who subscale of the Alien Quiz, two participants discontinued this subscale after two trials but continued with other measures. Following completion of the trials participants were requested to complete a visual analogue scale to assess the acceptability of the task.

Participants were then administered the remaining subscales of the Alien Quiz in the order Emotion Naming (free), Emotion Naming (cued), and Emotion Matching. Following completion of Emotion Matching task participants were requested to complete a visual analogue scale to assess the acceptability of the task. Finally, participants were asked which of the subscales they enjoyed the most.

Following completion of child participation their teacher was requested to complete the SDQ. These were left with the teacher and collected by the researcher one week after testing.

2.9. Data Analysis

2.9.1. Content Analysis

Questions from the Guess the Alien subtest of the Alien Quiz were recorded verbatim during the testing procedure and functioned as the primary data source for the content analysis. Each question asked by the participant was classified as a ‘coding unit’ for the analysis.

Previous research examining the utility of a “Guess Who” style or “20 Questions” tasks as a test of concept formation, evaluated the data source with regards to question quality, and relationship between questions (Alderson-Day & McGonigle-Chalmers, 2011; Pavitt, 2017). The current study is the first to evaluate this test format as a test of emotion recognition, as a result a different approach to content analysis was conducted. A two-phase content analysis was completed. The first phase aimed to identify the presence of stimuli characteristics in the data source. Stimuli characteristics were derived from categories identified in the creation of the test stimuli. A data driven, second phase of content analysis was completed next. This phase was initiated after the first five participants took part in the study and ended following review of the complete data set. On this basis some characteristics, which occurred less frequently, were combined into one code. Following the establishment of the final coding scheme, the data was coded by the researcher.

Content analysis relied on both quantitative and qualitative methodology, in keeping with Weber's (1990) suggested approach to content analysis. The two-phase content analysis derived codes qualitatively from both pre-established characteristics and the data source. This approach to the generation of codes allowed for data saturation to be achieved as all data fitted appropriately within the existing codes. The quantitative aspect of the content analysis involved conceptual analysis of the data source in order to determine the frequency of occurrence of questions about different characteristics in the stimuli, in order to explore the relative frequency of questions about emotion.

2.9.2. Additional Tests

As the Guess the Alien subtest is a novel approach to testing FER several quantitative measures were developed in order to establish useful methods of assessing FER ability in children.

2.9.2.1. *Use of emotion naming*: a ‘Use of Emotion Naming’ (UEN) score was calculated for each participant across all four trials of the Guess the Alien subtest. This is a categorical score which denotes whether participants referenced a trait coded as emotion in the content analysis. In creating this score only existence of

concept was coded for in the content analysis. UEN was developed to see whether there was a recognition of the presence of facial emotions in the stimuli across participants. This score is similar to emotion labelling tasks in established tests of FER. However, unlike traditional test methodology, this test did not explicitly ask the children about emotion, as a result this score could provide indication of children's ability to identify emotion stimuli in their day-to-day lives.

2.9.2.2. *Frequency of emotion naming:* a 'Frequency of Emotion Naming' (FEN) score was calculated for each participant. In scoring this measure, the researcher identified the number of times each participant referred to emotion naming in their responses across all four trials. This measure was established to determine whether children referred to emotions at similar frequencies in their responding. Similar to the UEN score this score aimed to establish a more ecologically valid measure of FER by assessing frequency of reference to FER without explicitly asking the child about emotion.

2.9.2.3. *Facial emotion matching:* the 'Facial Emotion Matching' (FEM) score was available only for participants who asked questions about the presence of a given emotion in the target alien. This score assessed the accuracy of participant's ability to identify other Aliens which that emotion within the game context. This score is similar to matching of emotion images methodology in existing tests of FER.

2.9.2.4. *Comparison to other measures:* Guess the Alien scores outlined above were compared to the traditional format Alien Quiz FER subtests, SDQ subtests and age using Spearman's Rank

2.9.3. Acceptability of the Task

Acceptability of the Guess the Alien subtest and Emotion Matching subtest were measured using a visual analogue scale depicting a 5 point Likert style scale and were compared using Wilcoxon Signed Rank.

2.10. Participants

2.10.1. Demographics

Participants were recruited from Year 3 and Year 5 of a London primary school. Twenty-five children took part in this study; 13 females and 12 males. Participant were aged between seven and ten years (89 to 124 months, M=107.71, SD=12.36). The majority of participants were from a White European background (80%, N=20). See Table 5 for a breakdown of participant ethnicity. Almost all participants were born in the UK (92%, N=23); just under half of participants (48%, N=12) had at least one British parent. Most participants spoke English as their primary language (80%, N=20); three participants reported that Polish was their primary language, two participants spoke Albanian as their primary language. All children who did not speak English as a primary language spoke English fluently; 11 participants described themselves as bilingual. One child reported that they had ongoing SLT support, one child described that they have hearing difficulties.

Table 5

Ethnicity of Participant Sample

Ethnicity	*n*
Albanian	3
White British	3
Polish	3
Other White Background	11
Black African	2
Mixed or Multiple Ethnic Background	2
Asian	1

2.10.2. Social and Behavioural Ability

SDQ scores were tested for normality using the Kolmogorov-Smirnov test with UK norms for five to ten year olds. The Kolmogorov-Smirnov test indicated that the Peer Problem SDQ score for the sample, $D(25)=1.740$, $p=.003$, significantly deviated from the normal distribution of this subscale in the UK population. The total group has

scores that are slightly higher with a wider standard deviation when compared to the UK norms.

Table 6

Descriptive Statistics for Age and SDQ Scores

	Total Group (*N*=25)		Year 3 (*n*=12)		Year 5 (*n*=13)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age in Months	107.71	12.35	95.64	3.69	118.86	3.54
SDQ Prosocial	6.96	2.56	6.75	3.13	7.15	1.99
SDQ Peer Problems	1.56	2.24	2.67	2.81	0.54	.66

Comparison of mean Peer Problem SDQ scores by year group (see Table 6) indicated that children in Year 3 had higher peer problem scores than those in Year 5; two Year 3 participants' Peer Problem scores fell into the very high category. The Kolmogorov-Smirnov test indicated that the Pro Social SDQ scores for the sample, $D(25) = .803$, $p = .489$, did not deviate from normal distribution. Compared to the UK population norms for the SDQ prosocial score the sample participants are similar to that of the UK population.

Table 7 displays descriptive statistics for the SDQ subscales. A boxplot of the SDQ Peer Problems subscale confirmed that the data was not normally distributed; two outliers and three extreme scores were identified. The skewness figure (2.109, SE .464) is above what is considered acceptable for normal distribution (Byrne, 2013; Hair, 2010). A Kolmogorov-Smirnov test indicated that this subscale was not normally distributed using at a significance value of 0.05.

Table 7

Descriptive Statistics and Normality Tests for SDQ Subscales

	Kolmogrov-Smirnov Z	Skewness (*SE* = .464)	Kurtosis (*SE* = .902)
SDQ Prosocial	.80	-.618	-.488
SDQ Peer Problems	1.74*	2.109	4.457

*p<0.05 indicating significant deviation from normality

No participant data was removed from the study, in order to allow for evaluation of how, children with a range of social competency, within a mainstream primary school performed on the Alien Quiz. Accordingly, non-parametric procedures were used when undertaking statistical analyses involving these variables.

3. RESULTS

This was an exploratory study and as such employed several methods of analyses to assess the research questions. For the Guess the Alien subtest, this analysis included exploring what features of the stimuli the children asked questions about; whether children adopted similar approaches to the game; how enjoyable children found the game; and whether it could be useful as a test of FER. For the Emotion Naming (free), Emotion Naming (cued) and Emotion Matching subtests analysis was completed to explore accuracy of identification of each emotion. For all subtests analysis was completed to explore whether test scores differ based on demographic characteristics and if there was a relationship between scores in these subtests and a real-world test of social functioning.

3.1. Exploratory and Qualitative Analysis of Game-like Task

Analysis of the game-like task, Guess the Alien, as a measure of FER was completed in a number of stages. Content analysis was first completed in order to understand how children used the game, and the types of questions they asked. Data emerging from the content analysis was then used to inform and evaluate the methods of scoring the test outlined in section 2.3.1.1.

3.1.1. Content Analysis

Development of categories in content analysis took place in two phases. In the first phase, categories outlined in the stimuli development scheme were used as coding categories. In phase two, remaining codes which did not fit into the original categories were analysed and were identified as containing key emotion traits (see section 2.4.1. for more information). This led to the creation of a new category ‘feature of emotion’. Three categories from phase one ‘horns’, ‘wings’, and ‘tail’ were combined into one larger category ‘appendage’ given the overall low number of codes for the three categories and the conceptual similarity of these categories. Table 8 provides the development of the two-phase scoring system with categories on the left-hand side representing those adapted from the predefined categories of

the stimuli development scheme. On the right categories which were added and removed following exploration of the data are outlined.

Table 8

Two Phase Content Analysis

Phase 1	Phase 2 Codes Added	Code Removed
Colour	Feature of emotion	Horns
Legs	Appendage	Wings
Emotion		Tail
Shape		
Horns		
Wings		
Tail		
Number		

The coding process led to a coding scheme with seven codes. Table 9 outlines the coding scheme with a description of the category, example of codes which meet the criteria for the category and examples of questions asked by participants which were allocated to this category. Most questions asked by participants matched the stimuli features included in the stimuli development scheme and were coded as such. For example, ‘colour’ is a category from the stimuli development scheme. Individual colours (red, yellow, blue, green, purple, grey, orange) were the “codes” used to identify the presence of the overarching category in the data and were also included in the stimuli development scheme. Some participants asked questions about the quality of the colour (ex. bright, dark). A new code (colour quality) was created for this question, and this code was included under the category of ‘colour’. The coding scheme was exhaustive; codes and their categories were mutually exclusive given that children were reminded that they could only ask about one feature at a time.

Table 9

Coding Scheme

Coding Category	Description of Category	Code example	Direct example
Colour	Questions relating to one of seven colours used in stimuli, Category includes used of questions which compile multiple colours together.	Red, Blue etc.	Is your alien blue? Is your alien a dark colour?
Legs	Questions relating to number of legs the target alien has.	One leg, no legs	Does your alien have more than two legs? Does your alien have five legs?
Emotion	Question relating to whether the target alien "is feeling" one of the six basic emotions. Child was required to name emotion or emotion synonym.	Happy, angry	Is your alien shocked? Is your alien afraid?
Feature of Emotion	Questions relating to a feature of one of the six basic emotions without naming an emotion. For example, "is the alien smiling"	Wide eyes, tongue out	Is your alien smiling? Is it's tongue sticking out?
Shape/ texture	Questions relating to one of three possible shapes or texture of the alien.	Round/ spotty Oval/furry	Is it furry? Is your alien round?
Appendage	Question relating to one of three appendages; horns, wings and a tail.	Horns, tail	Does your alien have wings? Does your alien have two horns?
Number	Question or pointing in order to question if specific alien is the target alien. Participants were requested not to ask other types of questions about numbers.	One, fifteen	Is it number five? Non verbal- points to image (number recorded by researcher)

3.1.1.1. *Quantitative Analysis:* following completion of coding and categorisation, trends of responding were explored (See table 10 below). Features of categories 'colour' and 'legs'' were created to have the same number of features as the 'emotion' category (see stimuli development scheme; Appendix A). This ensured that during game play use of questions in these categories would result in elimination of a similar number of features, i.e. would have similar strategic advantage for the participant to win the game. Some children did find ways to combine features to ask more strategic questions, for example with a question like "does the alien have more than two legs?". Although this method was used for a minority of questions it should be held in mind while comparing frequency of questions in each of these three categories.

Questions where participants identified and named an emotion accounted for only 7% of questions asked across trials; questions on emotion occurred at roughly a similar frequency across year groups. Eighteen of the 25 participants asked at least one question about emotion. An additional 4% of questions were coded to the 'features of emotions' category. Only one participant who did not ask about 'emotion' asked a question about a 'feature of emotion'. There were six participants who did not ask any questions in the 'emotion' and 'feature of emotion' category.

Table 10

Frequency and Percentages of Coding Categories Used Across All Trials

	Overall		Year 3		Year 5	
	Frequency	%	Frequency	%	Frequency	%
Colour	188	33%	109	40%	79	26%
Legs	116	20%	41	15%	75	25%
Emotion	40	7%	22	8%	18	6%
Feature of Emotion	25	4%	10	4%	10	3%
Shape	38	7%	7	3%	31	10%
Appendage	74	13%	35	13%	39	13%
Number	99	17%	47	17%	52	17%
Total	580		271		304	

Table 11

Means, Standard Deviation and Coefficient of Variation of Questions Asked Per Participant Across Three Categories

	Emotion	Colour	Legs
Mean	1.60	5.88	3.96
SD	1.41	4.20	3.30
CV	88.4%	55.7%	74.6%

Table 11 displays the mean, standard deviation and coefficient of variation (CV) of questions asked per participant in the three categories ‘colour’, ‘emotion’ and ‘legs’. All CV scores were high indicating that variation occurred in the frequency of questions on each category between participants, the ‘emotion’ category had the highest CV score indicating that the most variation occurred in this group. Overall ‘colour’ was the category most frequently asked about across trials; only one participant asked no questions about ‘colour’. ‘Legs’ was the second most frequent category identified across trials, 22 participants asked at least one question coded in this category. All children asked at least one question in the ‘number’ category; however this was most often to identify the target alien and as such it was unavoidable for each successful trial to have at least one question in this category.

Frequency of codes from the ‘emotion’ and ‘feature of emotion’ categories were analysed in order to understand which emotions were identified and named by participants. Table 12 below gives the number of times each emotion was named and the percentage of its occurrence. All of the six basic emotions were asked about at least once, though no child asked whether the alien had a neutral emotional expression. ‘Anger’ was the most frequently identified emotion and was identified by nine children. Thirteen children asked whether the alien was either ‘happy’ or ‘smiling’, given the complete overlap of aliens in these two codes it is possible that more children would have identified the emotion ‘joy’, but chose a more concrete feature of this emotion. Two children in Year 3 asked whether the alien “felt sick”, as this met the criteria for a synonym of ‘disgust’ it was coded as ‘disgust;’; however no child named ‘disgust’ spontaneously in Guess the Alien.

Table 12

Frequency and Percentage of Emotions and Features of Emotions Named Spontaneously in Guess the Alien

Emotions	Overall		*Year 3*		*Year 5*	
Joy	8	20%	3	14%	5	28%
Sadness	3	8%	1	5%	2	11%
Anger	12	30%	7	32%	5	28%
Fear	6	15%	4	18%	2	11%
Disgust	4	10%	4	18%	0	0%
Surprise	7	18%	3	14%	4	22%
Total	40	100%	22	55%	18	45%
Feature of Emotions						
Smiling	16	64%	4	40%	12	80%
Crying	1	4%	1	10%	0	0%
Tongue Out	5	20%	2	20%	3	20%
Mouth Open	1	4%	1	10%	0	0%
Mouth Closed	1	4%	1	10%	0	0%
Wide Eyes	1	4%	1	10%	0	0%
Total	25	100%	10	40%	15	60%

3.1.1.2. *Exploratory Content Analysis:* all attempted trials were successfully completed in under 10 questions and were considered to be successful attempts. Two children completed only two of the four trials of the Guess the Alien Quiz due to time constraints. All children used full sentences to ask questions about the aliens. The majority of questions fit into the pre-established coding scheme, identified during the allocation of features of the stimuli. This indicated that children were able to survey the character set in front of them and could group the characteristics into categories which allowed them to eliminate aliens and win the game. No child repeated a question twice in the same trial indicating that children were able to remember the questions they had already asked.

'Colour' and 'legs' were asked about at a higher frequency than 'emotion' despite all three having the same elimination potential (1/7 characters). Children were not asked about why they asked certain questions, as a result it is unclear why features with the same elimination potential were asked about at different rates. It is possible

that ‘colour’ and ‘legs’ are more concrete characteristics and were used more frequently as children found them more salient. This seems particularly likely as ‘shape’ and the individual ‘appendage’ (horn, tail, wings) had higher elimination potential (1/3 characters) but were asked about at a lower frequency than ‘colour’ and ‘legs’. ‘shape’ and ‘emotion’ categories were relied upon at similar frequencies.

3.2. Other Measures of the Game-Like Task

Alongside content analysis, measures developed from the naming and matching measures used in previous tests of FER, were also derived and evaluated for this study. Measures created to assess FER in the Guess the Alien Quiz were developed by combining all four trials of the game. This was due to low frequency and inconstant usage of questions relating to emotion by participants in the game. These measures were:

- Frequency of Emotion Naming (FEN)
- Use of Emotion Naming (UEN)
- Facial Emotion Matching (FEM)

3.2.1. Frequency of Emotion Naming and Use of Emotion Naming

The FEN score was calculated by counting the number of times the ‘emotion’ category was coded per participant across all trials. Unlike traditional tests of FER this measure aims to establish the ability of children to spontaneously identify and name emotions. Given the low FEN scores among participants the UEN score was developed to establish as a measure to assess use of affect naming. The UEN score differentiated between participants who used emotion as a category in Guess the Alien and those who did not use emotion at all. Table 13 displays participant scores across all measures of Guess the Alien.

3.2.2. Facial Emotion Matching (FEM)

The Guess the Alien FEM score was developed to measure participant accuracy in matching facial affect; it is based upon the matching of emotion images methodology in existing tests of FER. Given that the Guess the Alien subtest has a game-like methodology, a novel method for scoring this measure was developed. Scores were

calculated by using the participant record sheet to recreate the playing board based on the questions asked by the participant. Participants were awarded one point per correctly matched emotion, one point if they did not miss any appropriate matches and one point if they did not incorrectly match any characters. Participant scores were then divided by the total number of 'emotion' questions they asked in order to control for participants who asked more questions having absolutely greater but lower proportionally FEM scores.

Table 13

Mean Guess the Alien Scores

Participant	FEN score	UEN score	FEM score
1	1	Yes	0.00
2	3	Yes	3.67
3	0	No	-
4	1	Yes	4.00
5*	0	No	-
6	1	Yes	1.00
7	1	Yes	3.00
8	1	Yes	3.00
9	1	Yes	5.00
10	0	No	-
11	3	Yes	4.00
12	5	Yes	3.2
13	2	Yes	5.00
14	2	Yes	3.00
15	0	No	-
16	3	Yes	3.00
17	3	Yes	4.00
18	0	No	-
19	2	Yes	4.50
20	0	No	-
21	2	Yes	2.50
22	2	Yes	4.00
23	3	Yes	3.00
24	4	Yes	3.75
25*	0	No	-

* Participant did not complete all four trials.

FEM scores were calculated for the 18 participants who asked at least one question of FER. The maximum FEM score achieved was 5, the minimum score was 0. The mean score achieved by participants was 3.31.

3.3. Exploratory Analysis of Other Alien Quiz Measures of FER

3.3.1. Emotion Naming (Free)

Table 14 displays participant responses in the Emotion Naming (free) subtest. These were analysed by emotion to understand how accurate children were in identifying the emotion depicted by the cartoon alien. 'joy' and 'anger' emotion stimuli had accurate naming rates (above 90%) across age groups. Unexpectedly, accuracy of naming of 'fear' and 'surprise' decreased as children grew older; naming of 'disgust' improved with age. Identification of 'sadness' from the image was relatively low, with only 52% of participants correctly naming this.

3.3.2. Emotion Naming (Cued)

Accuracy in Emotion Naming (cued) subtest, which required participants to match the image to the correct emotion label, is shown in Table 15. Overall accuracy in labelling of 'joy', 'anger' and 'surprise' was the greatest among participants. Recognition of 'sadness' (88% correct) and 'disgust' (84% correct) also had high accuracy when a choice of labels was provided. Recognition of 'fear' was lowest among participants; six participants labelled 'fear' incorrectly.

3.3.3. Emotion Matching

Participant accuracy scores for correctly matching the target emotion with all corresponding emotions on the Alien Quiz grid is shown on Table 16. In keeping with the other subtests, accuracy was highest for 'joy' and 'anger'. Participants demonstrated poorer performance in matching accuracy for 'surprise' than in other subtests. Fatigue possibly played a role in performance as the 'surprise' in the emotion matching subtest was the final subtest administered in the test battery.

Table 14

Accuracy of Performance in Emotion Naming (Free) by Emotion

	Overall (*N*=25)	Year 3 (*n*=12)	Year 5 (*n*=13)
Joy	25 (100%)	12 (100%)	13 (100%)
Anger	23 (92%)	11 (92%)	12 (92%)
Disgust	10 (40%)	4 (33%)	6 (46%)
Sadness	13 (52%)	6 (50%)	7 (54%)
Fear	14 (56%)	10 (83%)	4 (31%)
Surprise	21 (84%)	11 (92%)	10 (77%)

Table 15

Accuracy of Performance in Emotion Naming (Cued) by Emotion

	Overall (*N*=25)	Year 3 (*n*=12)	Year 5 (*n*=13)
Joy	23 (92%)	10 (83%)	13 (100%)
Anger	24 (96%)	12 (100%)	12 (92%)
Disgust	21 (84%)	10 (83%)	11 (85%)
Sadness	22 (88%)	10 (83%)	12 (93%)
Fear	19 (76%)	8 (67%)	11 (85%)
Surprise	24 (96%)	11 (92%)	13 (100%)

Table 16

Accuracy of Performance in Emotion Matching by Emotion

	Overall (*N*=25)	Year 3 (*n*=12)	Year 5 (*n*=13)
Joy	22 (88%)	11 (92%)	13 (87%)
Anger	22 (88%)	10 (83%)	14 (93%)
Disgust	12 (48%)	5 (42%)	9 (63%)
Sadness	12 (48%)	5 (42%)	7 (54%)
Fear	15 (60%)	8 (67%)	7 (54%)
Surprise	10 (40%)	2 (17%)	7 (54%)

3.4. Methods of Quantitative Analysis

SPSS version 27 (IBM Corp, 2020) was used to analyse quantitative data. Exploratory data analysis was completed on all continuous variables in the data set in order to explore patterns in the data, errors, outliers and data distribution. Boxplots, skewness (< +/-1) and kurtosis (< +/- 3) were inspected; a summary of these findings are provided in Table 17.

Exploratory data analysis indicated that a many of the variables were not normally distributed. Due to the non-normal distribution, and the relatively small sample size, the data was analysed using nonparametric tests. Mann-Whitney U tests were used to evaluate whether mean scores on the Alien Quiz subtests differed based on demographic groups. Spearman's rank correlations were used to measure the direction and strengths of relationships between Alien Quiz subtests, SDQ measures and age. Cohen's (1988) guidelines were used to interpret the magnitude of correlations: r = .10 - .29 is a small effect size, r = .30 - .49 is a moderate effect size, and r > 0.50 is considered a large effect size.

Table 17

Summary of Exploratory Data Analysis

	N	Mean	SD	Min	Max	Skewness (SE)	Kurtosis (SE)
Age (months)	25	108.71	12.35	89.13	123.83	-0.09 (.46)	-1.82 (.90)
SDQ Prosocial	25	6.96	2.56	1	10	-0.62 (.46)	-0.49 (.90)
SDQ Peer Problems	25	1.56	2.24	0	9	2.11 (.46)	4.46 (.90)
GTA FEM	18	3.31	1.26	0	5	-1.23 (.54)	2.08 (1.04)
GTA FEN	25	1.60	1.41	0	5	0.60 (.46)	-0.30 (.90)
Emotion Naming (free)	25	4.24	1.09	2	6	-0.10 (.46)	-0.74 (.90)
Emotion Naming (cued)	25	5.32	1.03	2	6	-1.71 (.46)	3.11 (.90)
Emotion Matching	25	12.08	1.98	8	15	-0.30 (.46)	-0.99 (.90)
FER Total	25	21.64	2.94	12	26	-1.42 (.46)	3.56 (.90)

3.5. Performance on Measures of FER

3.5.1. Guess the Alien Subtest

Mann-Whitney U was used on the continuous Guess the Alien subtest measures, FEN and FEM, in order to explore whether there is a relationship between performance on these measures and demographic variables year group, sex and language group. Language group refers to English only speakers (monolingual) and English and additional language speakers (bilingual). A summary of Guess the Alien subtest means by year group, sex and language can be found on Table 18.

3.5.1.1. *Year Group:* somewhat unexpectedly participants in Year 3 performed better than Year 5 participants in both subtests; Mann-Whitney U tests indicated that this difference between groups was not substantive for both FEN *(U* = 66.00, *Z* = -0.67, *r* = .13, exact sig. = .525) and FEM *(U* = 39.00, *Z* = -0.09, *r* = .02, exact sig. = .951) scores. A small effect size was observed for FEN scores.

3.5.1.2. *Sex: s*imilar performance was observed across sex for FEN and FER scores. Mann-Whitney U tests confirmed no substantive difference in scores, there was a small effect size for both FEN *(U* = 70.00, *Z* = -0.45, *r* = .09, exact sig. = .673) and FEM *(U* = 35.50, *Z* = -0.40, *r* = .09, exact sig. = .710) scores.

3.5.1.3. *Bilingualism:* children who spoke only English had higher FEN scores, than pupils who spoke both English and an additional language. A Mann-Whitney U test indicated a medium effect size whereby participants who spoke only English asked more questions about emotions than those who were bilingual *(U* = 40.00, *Z* = -2.08, *r* = .42, exact sig = .038). No difference was observed between language groups in the FEM score *(U* = 38.00, *Z* = -.05, *r* = .01, exact sig = .987).

Table 18

Summary of Guess the Alien Means and SD by Year, Sex, and Language Group

	N	Total	Year 3	Year 5	Female	Male	Mono-lingual	Bi-lingual
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
GTA FEN	25	1.60 (1.41)	1.83 (1.59)	1.38 (1.26)	1.77 (1.59)	1.42 (1.24)	2.14 (1.56)	0.91 (1.28)
GTA FEM	18	3.31 (1.26)	3.56 (0.87)	3.14 (1.52)	3.30 (1.04)	3.33 (1.42)	3.28 (1.28)	3.36 (1.31)

3.5.2. Traditional FER Format Alien Quiz Subtests

Mann-Whitney U tests, exploring whether there is a relationship between performance on the three traditional format FER subtests of the Alien Quiz; emotion naming (free), emotion naming (cued), and emotion matching, and demographic variables year group, sex and bilingualism were completed.

3.5.2.1. *Year group:* participants in Year 5 performed better in the total FER score, cued and matching subtests, but not in the free naming subtest. Mann-Whitney U tests indicated that difference between year group was not substantive; see Table 19 for a summary of Mann-Whitney U test statistics, effect size and significance.

3.5.2.2. *Sex:* Table 20 provides a summary of traditional format Alien Quiz subtest Mann-Whitney U test results by sex. Female participants had higher mean rank scores on every traditional format FER subtest of the Alien Quiz. Mann-Whitney U tests indicated that group differences were not substantive, there was a small to medium effect size for all measures.

3.5.2.3. *Bilingualism:* Table 21 provides a summary of traditional format Alien Quiz subtest Mann-Whitney U test results by language group. Participants who spoke only English performed better than bilingual participants in all subtests using the traditional format. Mann-Whitney U tests indicated that difference between language groups was not substantive.

Table 19

Summary of Mann-Whitney U Test Results of Alien Quiz Subtest Scores for Year 3 and Year 5 Groups

Subtest	Year Group	*n*	Mean Rank	*U*	*Z*	*r*	exact sig.
Emotion Naming (free)	Year 3	12	14.92	55.00	-1.30	.26	.209
	Year 5	13	11.23				
Emotion Naming (cued)	Year 3	12	12.13	67.50	-0.65	.13	.563
	Year 5	13	13.81				
Emotion Matching	Year 3	12	10.88	52.50	-1.41	.28	.165
	Year 5	13	14.96				
FER Total Score	Year 3	12	12.17	68.00	-0.55	.11	.599
	Year 5	13	13.77				

Table 20

Summary of Mann-Whitney U Test Results of Alien Quiz Subtest Scores for Males and Females

Subtest	Sex	*n*	Mean Rank	*U*	*Z*	*r*	exact sig.
Emotion Naming (free)	Male	12	12.13	67.50	-0.59	.12	.583
	Female	13	13.81				
Emotion Naming (cued)	Male	12	10.83	52.00	-1.61	.32	.103
	Female	13	15.00				
Emotion Matching	Male	12	11.38	58.50	-1.08	.22	.296
	Female	13	14.50				
FER Total Score	Male	12	11.13	55.50	-1.24	.25	.225
	Female	13	14.73				

Table 21

Summary of Mann-Whitney U test Results of Alien Quiz Subtest Scores for Monolinguals and Bilinguals

Subtest	Group	*n*	Mean Rank	*U*	*Z*	*r*	exact sig.
Emotion Naming (free)	Bilingual	11	11.05	55.50	-1.22	.24	.233
	Monolingual	14	14.54				
Emotion Naming (cued)	Bilingual	11	12.91	76.00	-0.06	.01	.978
	Monolingual	14	13.07				
Emotion Matching	Bilingual	11	12.14	67.50	-0.53	.11	.615
	Monolingual	14	13.07				
FER Total Score	Bilingual	11	11.23	57.50	-1.08	.21	.291
	Monolingual	14	14.39				

3.6. Relationships Between Measures

Spearman's rank correlation was undertaken to explore the relationship between the Alien Quiz measures, age and SDQ subtest scores. Table 22 gives a summary of correlation coefficients. Age exhibited a moderate negative correlation with the SDQ peer problem scores, indicating that younger children had higher peer problem scores. SDQ prosocial and peer problem subscales exhibited no associations with measure of FER or each other.

The FEN score on the Guess the Alien subtest exhibited a moderate association with the Total FER score: greater frequency of naming emotions was associated with better total scores of FER. The FEM Guess the Alien score did not exhibit and associations with other measures.

Table 22

Spearman's Rank Correlation Comparing Alien Quiz Subtests with Age and SDQ

			SDQ		Guess the Alien		Adapted Traditional FER subtest			
		Age	Pro-social	Peer Problem	Facial Emotion Matching	Facial Emotion Naming	Emotion Naming (free)	Emotion Naming (cued)	Emotion Matching	Total
Age		1.000								
SDQ	Prosocial	.147	1.000							
	Peer Problems	***-.428**	-.200	1.000						
GTA	FEM	.048	.285	.159	1.000					
	FEN	.039	.155	.110	.165	1.000				
FER	EN (free)	-.102	-.316	.037	-.193	-.035	1.000			
	EN (cued)	*.**455**	.091	-.204	-.166	.199	*.**432**	1.000		
	Emotion Matching	*.**397**	.107	-.077	-.154	.114	-.103	.210	1.000	
	Total	.**389**	-.024	-.085	-.096	*.**409**	.442	.659	.756	1.000

Figure in bold indicate moderate to large effect size

* Indicates correlation is significant at 0.05 level (2-tailed)

** Indicates correlation is significant at 0.01 level (2-tailed)

Unsurprisingly, all scores of the adapted traditional FER subtests demonstrated moderate to high correlations with the Total FER score. Emotion Naming (free) and Emotion Naming (cued) exhibited a moderate correlation, indicating that better performance on Emotion Naming (free) was associated with better performance on Emotion Naming (cued). Emotion Naming (cued) and Emotion Matching had moderate correlations with age, indicating that older children performed better on these subtests. There was no association between Emotion Matching and other measures.

3.7. Acceptability of Task

A Wilcoxon Signed Rank Test was performed to determine if there was a substantive difference in the mean score of acceptability of task for the Guess the Alien subtest versus the Emotion Matching subtest. The test revealed that there was a significant difference in mean acceptability score between the two groups, with a large effect size (N = 24, Z = -2.88, r = .59, exact sig. = 0.006), indicating that participants found the Guess the Alien task more enjoyable than the matching task.

4. DISCUSSION

This section will revisit the research questions which this study addresses, exploring the results of the study in relation to these questions. Previous literature considered in the introduction will be reconsidered to contextualise the study findings. This section will also critically explore the strengths and weaknesses of the study with regards to study design and methods, alongside conceptual challenges associated with creating game-like tests. The conclusion of this section will offer an overview of the clinical implications of the study and potential directions for further research.

4.1. Summary of Findings

This was an exploratory, pilot study investigating the utility of a new social cognitive tool to measure facial emotion recognition (FER), the Alien Quiz. This tool comprises four subtests; the first task Guess the Alien applies a '20-Questions' style format as an indirect measure of FER. Novel cartoon facial emotion stimuli were created based upon Ekman and Friesen's (2003) established features of the six basic emotions. Unlike traditional measures of FER, facial emotions were depicted on novel aliens in order to create culturally neutral stimuli. Children reported that they enjoyed the game-like task more than the more traditional FER task formats and were observed to engage well with all measures. All children were able to complete the task successfully by guessing the target alien. Most children used facial emotion as a relevant feature of the alien stimuli, although at a lower rate than other features presented with the same frequency. Content analysis of the game-like task lead to the development of two novel scoring systems which may have utility as a measure of FER. The three remaining subtests use the novel emotion stimuli in formats similar to traditional tests of FER. Children were observed to engage well with these tasks; emotion recognition accuracy using these measures was similar to that reported in the existing literature.

4.2. Research Questions

This study addressed the research questions below to explore the feasibility and utility of a novel test of FER. In particular, this study focused on the utility of the incorporation of a game-like procedure in the Guess the Alien subtest. A similar format, using a '20-questions' style approach, has been used in tests of concept formation (Alderson-Day & McGonigle-Chalmers, 2011; Pavitt, 2017), however to my knowledge this is the first time such an approach has been adopted in a test of FER.

This study aims to address the following research questions:

- Do typically developing children detect and draw upon emotional expression as a means of categorisation in the Guess the Alien subtest?
- Do typically developing children apply similar strategies during testing with the Alien Quiz?
- Do primary school aged children engage well with Guess the Alien as a test of FER?
- Is there a relationship between performance on the Alien Quiz and a test of real-world social functioning?

4.2.1. Research Question 1

Content, descriptive, and non-parametric analyses were utilised in order to explore whether children detected and drew upon emotional expression as a means of categorisation in Guess the Alien. Questions asked by participants in the Guess the Alien subtest were transcribed by the researcher and analysed using a two-phase content analysis. Codes were categorised during content analysis and quantitative descriptive analysis was undertaken to understand the strategies used by children during the task, and the frequency of their reliance upon emotion expression as a means of categorisation. Novel scoring measures of FER were developed to investigate whether use of categorisation of emotion in Guess the Alien could be used to assess FER ability. Finally, hypothesis testing was completed in order to explore whether categorisation of emotions was related to year, sex or language group of participants.

4.2.1.1. *Interpretation of findings:* The results partially supported the hypothesis that children would detect and draw upon 'emotional expression' as a means of categorisation in Guess the Alien. Most children asked questions involving facial emotions at least once across trials of Guess the Alien; indicating that most children identified the presence of emotion stimuli in the test material and were able to apply it as part of their strategy in the game. Despite this, children relied upon the 'emotion' category less than they asked about Alien's 'colour' and 'legs' which, having the same elimination potential (1/7), afforded a similar strategic advantage in the game. It is possible that children asked questions from categories which relied upon developmental skills established prior to primary school age, such as colour naming (Bornstein, 1985; Pitchford & Mullen, 2001) and counting (Wynn, 1990), than emotion identification which emerges throughout primary school age (Chronaki et al., 2015; Kujawa et al., 2014; Lawrence et al., 2015). Given the high number of questions that could be asked in each category (seven) children were able to rely upon questions based on these other categories to successfully complete the trial.

'Colour' and 'legs' accounted for more than half of all questions asked by participants. This was somewhat surprising given that these categories have a lower elimination potential (1/7) than questions falling in the 'shape' and 'appendage' categories, which have 1/3 elimination potential. While a "lucky guess" on a low elimination question could result in a positive outcome in the game, the more strategic and reliable approach would be to ask questions with better elimination potential. This indicates that perhaps the 'colour' and 'legs' categories were more salient to the children participating. It is notable that participants inquired about the alien's 'shape' at a similar frequency to questions about 'emotion'. 'Shape' was designed as a more complex category where two features combined to create the category code (for example, a participant could ask if the alien was round or had spots, but all round aliens had spots and vice versa). Similarly, in the 'emotion' category, eye and mouth features were often combined in order to created individual emotion codes (for example, 'happy' has narrowed eyes and an up turned mouth). It is possible that 'emotion' and 'shape' may have been perceived as more complex attributes, and so more challenging to identify and match with other characters.

The content analysis indicating that most children detected and drew upon emotional expression as a means of categorisation, was used to inform the development of novel scoring methods in the Guess the Alien subtest. Two scoring measures were developed: Frequency of Emotion Naming (FEN), which assessed participant ability to spontaneously identify and name emotional expressions, and Facial Emotion Matching (FEM), which assessed participant accuracy in matching facial emotion expression. Higher scores on these measures may indicate more accurate identification, naming, and matching of facial expressions.

The relatively low number of questions in the 'emotion' category resulted in FER scoring limitations. A strength of the Guess the Alien design is the combination of emotion identification, naming, and matching within a game-like format. This, however, is limited by the potential for children to select other categories, and not engage with the FER components of the test. While failure to engage in FER may tell us something about the child's FER ability, it results in challenges for scoring of the task. Despite these methodological, limitations a moderate association was found between the Frequency of Emotion Naming score in Guess the Alien and the Total FER score (which is the combined score of adapted traditional measures of FER). No association was found between the Guess the Alien scores and individual FER subscales. This may be representative of Guess the Alien requiring a combination of different aspects of FER.

Challenges in scoring also emerged when considering the matching component of the task. The matching format used in traditional tests of FER requires the participant to match the target emotion stimuli to one of a choice of four to six other emotion stimuli displaying the same emotion. In the Guess the Alien task, participants did not have any constraints on how many other emotions they could match to. This resulted in large scores as some children would correctly match some emotion stimuli but incorrectly match up to three other emotions leading to difficulties in establishing a scoring criterion. No associations were found between the Guess the Alien FEM score and other measures of FER included in this study. It is possible that a game format, which employed more restrictions around the number of items that can be matched per move, might ameliorate the difficulties in assessing this ability. Given

the exploratory nature of this study it was important to attempt to assess the utility of different measures, however it appears that, with the existing game format, FEM may not be a useful measure of FER.

Analysis of emotion naming in the Guess the Alien subtest, by demographic variables (year, sex and language group) indicated that Year 3 participants and females performed better on this measure. While these differences were not substantive it is interesting that younger children appeared to engage more with the emotion stimuli and may indicate that this test is more appropriate for younger children.

A difference in emotion naming was found whereby participants who spoke only English demonstrated better performance than those who spoke English and an additional language. It is not known why bilingual English speakers drew upon emotional expression as a means of categorisation, at a lower frequency than monolingual English speakers. One possible explanation for this, provided by Kazanas et al., (2019), is that the first language acquired by bilinguals retains an emotion processing advantage. Given that most study participants acquired English as a secondary language through English medium education it seems a plausible explanation. Future research could further attempt to understand these differences. Bilingual participants also had lower scores in the other FER subtests. These differences were largest in the least directive measures (Guess the Alien and Free Emotion Naming) and smaller in the more directive, Emotion Naming (cued), subtest.

4.2.2. Research Question 2

Participant responses and scores on all subtests of the Alien Quiz underwent content, descriptive and non-parametric analysis to determine whether typically developing children applied similar strategies in the Alien Quiz.

4.2.2.1. *Interpretation of Guess the Alien results:* content and descriptive analysis results indicated that children from non-clinical samples use similar strategies when playing Guess the Alien. All children appeared to understand and adhere to the rules of the game, regardless of whether they had played a ‘20-Questions’ style game

before. Children asked, in complete sentences, questions about the stimuli characteristics in a manner which attributes were placed under superordinate categories. For the most part, categories of grouping used by the children, aligned with the researcher's pre-established characteristics. All children completed each trial using fewer than 10 questions, the cut-off point indicating a successful attempt.

4.2.2.2. *Interpretation of findings- adaptation of traditional measures of FER:* Descriptive and non-parametric analysis was undertaken on the subtests which used the traditional test of FER format. Results from these subtests indicated that children successfully engaged with Emotion Matching, Free, and Cued Emotion Naming tasks which used culturally neutral cartoon facial emotion stimuli. Accuracy of Emotion Naming (cued) in the Alien Quiz (the most commonly used format in tests of FER) was on par with, or better than, those reported by Lawrence et al. (2015) in their study investigating the development of FER among school children, using the Ekman-Friesen Pictures of Facial Affect test (Ekman & Friesen, 1976).

Accuracy of FER varied by subtest. Accuracy of naming in the Emotion Naming (cued) subtest varied from 76% correct for 'fear' to 96% correct for 'anger' and 'surprise'. In contrast correct performance in the Emotion Naming (free) subtest ranged from 40% for 'disgust' to 100% for 'joy'. Poorer performance on free naming is somewhat expected given that participants were required to spontaneously name emotion words, as such it is a more developmentally challenging task than matching forced choice to predefined labels. Emotion Naming (free) accuracy of 'sadness' was lower than expected; just over half of participants correctly named this emotion. Given the relatively early acquisition of emotion naming of 'sadness', at four to six years (Baron-Cohen et al., 2010), it is possible that there are issues with the 'sadness' emotion stimuli.

Previous literature indicating that naming of some emotions (such as 'anger', 'surprise', and 'fear') consolidates around seven to eight years would suggest that younger children might find Emotion Naming (free) a more challenging task. Interestingly, younger children performed better in this subtest than the older group. The opposite was true for the other traditional format subtests.

Monolingual English speakers performed better on all FER measures than their peers who spoke English and an additional language. This difference was greatest in the free naming subtest; possibly because bilingual participants may have found generating emotion terms more challenging.

4.2.3. Research Question 3

This research question investigated the acceptability of the game-like Guess the Alien subtest in comparison to a traditional format FER task. To assess this question, participant Likert scales which measured enjoyment of Guess the Alien and the Emotion Matching subtests were compared.

4.2.3.1. *Interpretation of the findings:* non-parametric analysis supported the hypothesis that children would enjoy the Guess the Alien subtest more than a traditional format FER task (Emotion Matching). Observation indicated that children engaged well the game-like subtest and appeared motivated to do well in it. This is particularly notable considering that the Guess the Alien subtest was longer (composed of four trials), in comparison to the other subtests of only one trial. Despite the repeated effort children preferred the Guess the Alien format.

4.2.4. Research Question 4

The final research question addressed whether a relationship existed between real word social functioning, as measured by peer problems and pro social SDQ subscales, and measures of FER.

4.2.4.1. *Interpretation of findings:* No association was found between measures of FER on the Alien Quiz and SDQ subscales. Although the Alien Quiz does not aim to measure the child's overall social and emotion functioning, previous literature has found positive associations between FER ability and SDQ scores, with improved performance on FER associated with fewer social and emotional difficulties (Staff et al., 2022; Wells et al., 2020). Given the increasing emphasis that tests of cognition can also be translated into indications of real world functioning (Gioia & Isquith, 2011), evaluation of predictive validity alongside construct validity is important in the development of new test. It was hypothesised that an association would exist

between measures of FER and the SDQ subscales. However this was not the case for this study.

In their comprehensive chapter on social cognition in children, Beaudoin and Beauchamp (2020) caution that social cognition ability cannot always be inferred from everyday behaviour: while poor FER and may result in poor social behaviour, it is also possible that poor social behaviour is due to other factors such as mood or motivation. Thus, consideration of other factors which may impact upon social and emotional functioning may contribute to understand the lack of association found in this study. Alternatively another measure of real life social cognition may be more appropriate. Leppänen and Hietanen (2001) found that social adjustment, assessed through a teacher reported measure, and peer popularity, was reliably related to FER ability for girls.

4.3. Relationship to Previous Research

The ability to interpret facial expression of emotion supports successful social interactions and interpersonal functioning in social groups (Schultz, 2005). Deficits in FER have been associated with diverse aetiologies including brain injury, neurodevelopmental disorders (Schmidt et al., 2010), mental health conditions (Collin et al., 2013), and social and environmental causes (Kujawa et al., 2014). The impact of FER deficits have been associated with interpersonal (Collin et al., 2013), behavioural (Castro et al., 2010), and emotional (de la Torre-Luque et al., 2022) difficulties. Despite the importance of FER in social and emotion development in children, few tests of FER have been designed specifically for children. Those that have, rely upon the same format as tests designed for adults despite difference in how children and adults engage with psychological assessment. Existing tests use photographs of adult facial stimuli, which although suitable for testing of adults given face validity enhances adult motivation to complete testing (Nevo, 1985), may not promote engagement for children. In psychological testing of children, use of game-like tasks has been found instead to enhance engagement (Lumsden et al., 2016; Song et al., 2020).

This research aimed to create and assess the feasibility of a novel, game-like test of FER designed specifically for children, with their developmental needs and abilities in mind. Influenced by previous research investigating the use of a '20-Question' style task in the assessment of concept formation (Alderson-Day & McGonigle-Chalmers, 2011; Pavitt, 2017), this research focused on developing a test of FER which had the following features:

- Novel, child friendly facial emotion stimuli.
- A game-like structure, using a commonly used format in children's games.
- Enjoyable and engaging for children.
- Developmentally appropriate for children, with easy to comprehend instructions.
- Sensitivity to measure FER ability.
- Elicited typical patterns of responding in typically developing children which allow for the development of standardised FER scoring.

4.3.1. Game-like Procedure

To my knowledge, the Guess the Alien subtest of the Alien Quiz, is first time a game-like procedure has been used as a test of FER for children. In keeping with previous research findings that a game-like format enhances engagement in and enjoyment of cognitive tests (Hawkins et al., 2013; Lumsden et al., 2016; Song et al., 2020), participants expressed a preference for Guess the Alien in contrast to the more traditional format of FER ability. Guess the Alien incorporated using regular positive feedback such as "you're doing really well" throughout the game and "well done, you've won the game" after successful completion in order to promoting the 'reward of the task', an important mechanism through which game-like tasks enhance motivation (Lewis et al., 2016). Participants appeared to respond well to this feedback; however it is possible that a two-player format, whereby the participant engages in competition with another player would further enhance motivation on the task.

Attempts were made to support the broad range of cognitive functions which a game-like task might rely upon (Granic et al., 2014). Working memory demands were reduced by encouraging children to turn over the cards of Aliens which their

questions had eliminated (Alderson-Day & McGonigle-Chalmers, 2011). Four short trials of Guess the Alien were administered alongside one trial each of the other measures in order to support sustained attention, which undergoes rapid development in children up to ten years (Betts et al., 2006). A '20-Questions' style task was employed as it is developmentally appropriate for the intended 6-11 age range of this study and was already a familiar game for many of the participants. Using a developmentally appropriate format for testing FER ensured that performance on the task was not limited by the developmental level of the child (Beaudoin & Beauchamp, 2020). Nonetheless, findings from this study suggest that bilingual children relied less on emotion categorisation than their monolingual peers and that linguistic ability may impact ability to engage with this task.

4.3.2. Novel Stimuli

Novel, cartoon stimuli were used in order to both appeal to children and support motivation and engagement with the tasks, and to enhance cultural neutrality of the stimuli. Previous research has indicated that an in-group advantage exists for the recognition of emotions across culture (Elfenbein & Ambady, 2002). Despite this in-group advantage, existing tests of FER give little consideration to the impact of cultural difference in the design of tests. Children in this study demonstrated similar accuracy levels in the Emotion Naming (cued) subtest to those documented by children in previous research (Lawrence et al., 2015); thus indicating that facial emotion depictions were adequately representative of their target. In keeping with the established development of FER ability (Beaudoin & Beauchamp, 2020) a moderate positive correlation was found between age and cued FER naming and FER matching ability, indicating that the novel cartoon stimuli remained recognisable for older children.

4.3.3. Comparison to Existing Measures

4.3.3.1. *Emotions included in FER test:* the majority of tests of FER investigate accuracy in identifying the six basic emotions established by Ekman and Friesen (2003). Incorporating the six basic emotions into the Guess the Alien subtest led to increased number of characteristics which the children could name and likely contributed to the reliance on two categories ('legs and 'colour') for over half the

questions asked. In particular as the categories ‘legs and ‘colour’ had the same number of characteristics in order to accurately compare the frequency of the ‘emotion’ category. The use of a subsection of the six basic emotions, for example, those that are established at an earlier age such as ‘joy’, ‘sadness’, ‘anger’ and ‘fear’, and a reduction of other characteristics may address these methodological difficulties and encourage more ‘emotion’ based questions.

4.3.3.2 *Scoring system:* Most established FER tasks, measure the number of times participants correctly identify the facial emotion. Each of the six basic emotions are displayed repeatedly with different actors portraying the emotion. As this was an exploratory pilot study, with several subtests, participants were only shown each emotion once in the Emotion Naming (free and cued) tasks, to reduce testing time and burden. In order to create a more valid and reliable task from these subtests participants could be shown the emotion stimuli of all 24 alien characters, in line with the format of existing tests. Unlike established FER matching tasks, this study asked participants to match the target emotion to “all the aliens who feel the same way”. This resulted in a scoring system which allocated a up to three points per emotion as opposed to established measures which simply allocated a point if participants correctly match the emotion.

The use of a game-like format for the Guess the Alien subtest required the creation of a novel scoring criteria. As described above, limitations in the design of the test may have contributed to limited usage of emotion stimuli. Notwithstanding this limitation, the Frequency of Emotion Naming score developed was found to have a moderate association with the total FER score indicating they likely measure related constructs.

4.3.3.3. *FER emotion matching format:* in creating matching scores for both the Guess the Alien and FER matching subtests it is possible that the testing format used may have relied upon skills other than FER such as visual-motor coordination, short term memory and inhibition. Existing FER matching tasks ask the participant to match the target emotion to one of four options (Paiva-Silva et al., 2016). The use of the full character set for this task in the Alien Quiz is a different approach. It is

possible that fewer characteristics and stricter rules (for example “you can only turn down four cards after each question”), might improve performance and scoring.

4.3.4. Participant Group

This exploratory study investigated the utility of the Alien Quiz with a sample of children from a mainstream primary school. Although no exclusion criteria were employed to restrict neurodiverse children or those with a learning disability from taking part, no children reported significant additional needs or having one-to-one support from a teaching assistant. It remains to be seen how children in clinical groups might engage with this task. Given the poorer performance on Frequency of Emotion Naming by bilingual participants, there is a possibility that good English language ability is a prerequisite for adequate performance, and that low emotion naming may not be indicative of poor FER ability for this group. In ‘20-Questions’ style measures of concept formation, children with ASD (Alderson-Day & McGonigle-Chalmers, 2011) and those with prefrontal cortex damage (Baldo et al., 2004) asked more questions than control groups. Research is required to understand the utility of the Alien Quiz with different clinical groups.

4.4. Critical Evaluation

4.4.1. Strengths of Current Study

4.4.1.1. *Study design:* To my knowledge this is the first study to explore the utility of a game-like format as a test of FER. This study makes a unique contribution by investigating the efficacy of a game-like structure, and the use of novel, culturally neutral stimuli with children in a mainstream school. The use of an exploratory design enabled analysis of children’s typical patterns of responding during the task, a feature of the design which was critical in understanding how children identified and named emotion stimuli. The design of the study also allowed for a preliminary understanding of whether the novel FER measure had the potential for the development of scoring based upon normative patterns of responding. Children were observed to employ similar strategies in Guess the Alien, indicating that in future research it may be possible to derive ‘norms’ for the task.

A strength of this study was the application of novel emotion stimuli in multiple testing formats, creating the four subtests of the Alien Quiz. Given the exploratory nature of this study, traditional formats of FER testing were employed alongside a game-like measure. This allowed for investigation into the efficacy of the novel stimuli in traditional testing format, and also comparison between the novel game-like format and traditional testing approaches.

The use of cartoon alien stimuli aimed to promote the cultural fairness in this task, given that a within-ethnicity group bias exists in FER. Traditional tests of FER primarily rely upon photographs of facial emotions, often depicted by white actors, potentially resulting in poorer performance by participants from ethnicities not represented in the stimuli. Given that most participants in this study were from a White European background it was not possible to address whether this approach was effective in creating a culturally fair test.

This task includes multiple formats of FER assessment and encompasses both indirect and direct measures. Unlike traditional tests of FER, Guess the Alien did not involve explicitly asking the children about facial emotion recognition. Like real world use of FER, children were assessed based upon their ability to notice and rely upon facial emotion stimuli. It was hypothesised that this element of the design may improve the predictive and ecological validity of the task.

The Alien Quiz also differed from traditional tests of FER in that the task was designed with the developmental level of children aged six to 11 years in mind. Traditional tests of FER have typically relied upon measures that were designed and standardised for adult participants, often with minor adaptions for children. Colourful, engaging facial emotion stimuli were created to appeal to this age group. The game-like format was adopted for the Guess the Alien task, and the emotion matching tasks was adapted in order to be more engaging than the traditional matching format. Language used during testing was standardised to be accessible for the youngest participants. It was anticipated that the use of child friendly stimuli and a game-like format would enhance child engagement and motivation which was

confirmed by high acceptability ratings by participants and high observed levels of engagement.

4.4.1.2. *Analysis:* the use of multiple methods of analysis, including content, descriptive and non-parametric analysis, was a strength of this research. A detailed two-phase content analysis was employed to understand the patterns of responding and use of emotion categorisation in the Guess the Alien task. Data were analysed descriptively and in comparison to each other, and demographic variables, using non-parametric assessment. Content analysis allowed for an understanding of how the test may be improved through recognition of characteristics which children responded to more frequently. Descriptive and non-parametric measures provided further insights into the usage of the Alien Quiz as a test of FER and demographic variables. Giving insight into how the test might be refined in future research.

4.4.2. Limitations of Current Study

4.4.2.1. *Participant sample:* all participants attended the same suburban London primary school resulting in a more homogenous sample than is representative of London as a whole. Exploratory analysis of the SDQ peer problems subscale found that participant scores differed from the UK norms, indicating that findings may not be generalisable to the UK population.

This research was also limited by the relatively small sample size (N=25). While adequate for understanding the patterns of responding in the task, this sample size limited the conclusions that can be drawn through statistical procedures other than non-parametric analysis. Despite efforts to recruit additional schools, limitations on the researcher's time, and school's resources, prevented this from occurring. Although the participant sample was adequate given that the focus of this study was test utility and development, homogeneity of participants did limit exploration of factors such as culture. Logistics in data collection resulted in only two-year groups (Year 3 and Year 5) participating in the study. As a result, no six year olds, or 11 year olds participated in the study. Inclusion of these ages would have been useful in more fully exploring the utility of the test for younger and older participants, and relationships with age.

4.4.2.2. *Design:* Guess the Alien and Emotion Naming (free) subtests required children to respond verbally. Attempts were made to reduce language demands by keeping verbal instructions simple and encouraging pointing. Despite this, non-parametric analysis suggested that the bilingual group relied less on the emotion category, and therefore that children who spoke another language found it more challenging to use emotion words. The inclusion of a verbal ability measure would have provided more certainty on the conclusions that could be drawn.

Although data was collected in a familiar setting, the room used for testing was not one that was familiar to the majority of the children. The use of this unfamiliar room, during a period when children would be typically in class, may have increased the novelty of the situation thus reducing the ecological validity of the task. Some participants asked questions about equipment in the room and appeared eager to explore the space. Obradović and colleagues (2018) suggest a possible alternative to remediate this: they propose that testing completed in the classroom with teachers or teaching assistants, would provide a more precise and ecologically valid results.

4.4.2.3. *The Alien Quiz limitations:* despite piloting the emotion stimuli on both adults and children, findings suggest that more extensive child piloting would be beneficial. Only two children, known to the researcher, were asked to provide feedback in the original piloting. It is possible that, knowing that the aliens were created by the researcher, the children were reluctant to give constructive criticism which may have improved the accuracy of emotions for children. Low accuracy in free naming of ‘sadness’ indicates that the representation of this emotion may benefit from review and further piloting. While the use of cartoon drawings greatly helped in the creation of culturally fair stimuli, further comparison to photographic emotion stimuli is required to ensure adequate validity (Paiva-Silva et al., 2016).

Finally, the game-like procedure introduced an element of chance to the format which led to limitations in scoring Guess the Alien. Although efforts were made to control for this by creating category characteristics which occurred at the same frequencies (for example ‘legs’, ‘colour’ and ‘emotion’), a “lucky guess” question

could still play a role in determining how many questions a participant had to ask. The large number of category characteristics also limited the design of the study as children did not have to rely on a variety of categories to win the game. Further research exploring the utility of Guess the Alien with fewer categories and characteristics is recommended.

4.5. Study Implications

4.5.1. FER Testing Implications

This study provides two unique contributions to current methods applied to evaluate FER.

4.5.1.1. *Game-like format:* The current study found that children rely upon emotion categorisation when introduced in a game-like format, and that frequency of emotion categorisation appears to have an association with other more traditional methods of measuring FER. Given the exploratory nature of this study, further conclusions cannot be drawn about the utility of a game-like format, however, findings indicate that, with adaptions, Guess the Alien may be a useful addition to the current battery of tests of FER, especially as all other tests of rely upon naming, labelling and matching formats originally designed for adults (Paiva-Silva et al., 2016).

4.5.1.2*. Culturally neutral stimuli:* The finding that children accurately labelled the stimuli in this study, at a similar or better rate to accuracy using photographic stimuli (Lawrence et al., 2015), indicates that non-human cartoon stimuli may be an appropriate alternative to human faces. This finding is of particular importance given the relative failure of existing tests of FER to control for the possible confounding influence of in-group advantage in recognising emotions across culture (Elfenbein & Ambady, 2002).

4.5.2. Clinical Implications

4.5.2.1*. Developmentally appropriate task*: evaluation of the current literature base on tests of FER indicates reliance on formats of testing of FER which were adapted from tests designed for adults and as a result do not have children's developmental

needs in mind. This task was designed specifically for children in the 6-11 age range, a period during which FER ability consolidates (Beaudoin & Beauchamp, 2020). The finding that children adopted similar strategies in naming and identifying emotions in a game-like test of FER indicates that in the future it may be possible to establish test norms and thus develop this task into a standardized measure of FER. The creation of an FER task designed to be developmentally appropriate for children will be beneficial for children who experience difficulties in FER and their parents as timely understanding of social cognitive difficulties may support prevention of further behavioural and emotional difficulties as children reach adolescence (Emond et al., 2007). Further development of this test of FER will also benefit those working with children with FER difficulties in both research and clinical practice.

The Alien Quiz aims to improve clinical utility through the child friendly, and engaging design of the task. Findings from this study indicate that the participants enjoyed completing the task, in particular the game-like component. Increased engagement and motivation properties of the task may be beneficial in engaging children who experience difficulty in sustained attention such as children with ADHD (Swaab-barneveld et al., 2000), autism spectrum disorder (Garretson et al., 1990; Vivanti et al., 2017), both groups with documented difficulties in FER (Airdrie et al., 2018; Collin et al., 2013; Dawson et al., 2002; Schultz, 2005).

The Alien Quiz is also relatively quick to administer, in part due to the short, child friendly instructions, making completion less time consuming for both children and administrators. As a result, this task may be less susceptible to discontinuation due to fatigue, a common occurrence in children with brain injuries (Gagner et al., 2015), another group with documented difficulties in FER (Schmidt et al., 2010; Snodgrass & Knott, 2006; Tonks et al., 2007a). Further, shorted testing time, child friendly materials, and high enjoyment ratings indicate that this task may be less anxiety provoking for children taking part. This could be of particular benefit in clinical practice, in particular in child mental health settings.

5.4.2.2. *Culturally neutral emotion stimuli:* the development of culturally neutral emotion stimuli on laminated cards has the advantage of flexibility of clinical utility.

From my experience of working with children with autism in CAMHs settings there can be challenges in finding emotion stimuli to support facial emotion recognition, for clients from diverse backgrounds, due to available resources predominantly depicting White facial emotion representations. The emotion stimuli for this task are intended to be culturally neutral; further research could clarify whether this results in increased accessibility and engagement for children from all backgrounds. The individual laminated card design of the emotion stimuli also means it can be applied flexibly for both FER testing and interventions to support FER ability.

4.6. Future Research Directions

This study was the initial phase of development of the Alien Quiz test of FER; future research could support the development of a standardised version of this task. This section will describe how future research could refine and develop the Alien Quiz. This section will also describe how future research could work towards the development of norms, for this test and also the investigate validity and reliability of the task.

4.6.1. Future Development of the Alien Quiz

To further develop the Alien Quiz as a measure of FER, amendments to some aspects of the stimuli are required. Findings from the Emotion Naming (free) subtest revealed a lower-than-expected level of accuracy in naming of ‘sadness’; just over half of participants correctly identified the emotion. Given that accuracy in naming of ‘joy’ and ‘sadness’ emerges between four and six years (Baron-Cohen et al., 2010), it should be the case that ‘sadness’ naming was established in the participants of this study. Further, accuracy in free naming of ‘joy’ was 100% indicating that there was no general difficulty in emotion naming within this sample. Sadness was also the emotion named at the lowest frequency in Guess the Alien, further indicating issues with the item.

It is recommended that additional piloting and feedback be completed for all emotions to ensure stimuli validity in future research. Expansion of emotion item piloting could be one way to ensure the accuracy of each emotion. Children could be

shown multiple variations of each emotion and asked to rank them based on how accurately they feel the images recommend the target emotion. At this stage children could also be asked for feedback on what they feel would improve the depiction of the emotion in the image, to allow the researcher to further refine the top ranked emotion image.

It is recommended that the emotion stimuli used in this study be edited to reduce the number of characteristics which children can inquire about in Guess the Alien. A reduction in characteristics would make the emotion stimuli more salient to the player and would lead to emotion discrimination becoming a more integral part of the game play. To maintain the ability to compare categorisation of emotion to other coding categories, reduction in emotions present in the game is also recommended. Given the importance of ensuring the developmental appropriateness of the task, the removal of ‘fear’ and ‘disgust’ from the emotion set is recommended. This is due to the continued refinement of ability to identify ‘fear’ and ‘disgust’ into late teenage years (Lawrence et al., 2015), whereas ‘joy', ‘sadness’, ‘anger’ and ‘surprise’ which are consistently recognisable in middle childhood (Kujawa et al., 2014; Lawrence et al., 2015).

Further exploration of the demographic and cognitive factors which influence performance on the Alien Quiz should be considered in order to address whether performance on the task depends on other developmental skills. The finding that monolingual English speakers performed better than bilingual peers suggests that verbal ability in a confounding feature. In future research direct recruitment of monolingual and bilingual participants in order to compare the performance of each group on this task would be an effective way to understand the contribution of language to performance. Additionally, inclusion of a measure of verbal ability might offer control for the possibility that perhaps culture or another demographic variables, accounts for the different performance between language groups.

4.6.2. Reliability

In order to determine the reliability of the Alien Quiz, the task could be administered to a larger sample size including children aged six and 11 who were not included in

this study due to logistical restrictions in data collection. Inclusion of participants from multiple schools in different geographical locations would help to ensure a representative sample is recruited. Efforts to prevent possible confounding variables should be considered such as by ensuring testing is completed in a similar location across schools (for example, a quiet room that is familiar to participants), participants are selected in a similar way (for example, entire class group as opposed to teacher selected participants). Increased sample size would support the statistical power of the study to detect relationships between test scores, demographic variables, and real world everyday social functioning. A test-retest study involving the participants who took part in this study would also a useful measure of reliability, of course practice effects should be considered with this approach.

Inter-item consistency of each trial of Guess the Alien should be assessed using Cronbach's alpha to investigate that each Guess the Alien trial is assessing the same aspect of FER. Inter-subtest correlations should also be completed in order to determine subtest redundancy: whether trials of the Alien Quiz are evaluating the same ability (R. J. Cohen et al., 1996). Piedmont (2014) recommends that correlation below .20 may not measure the same construct and that those above .40 may only measure a narrow portion of the construct in question. Future correlational analysis is required to understand whether further trials could be added or removed from the Alien Quiz.

4.6.3. Validity

Future research should assess for the concurrent validity of the Alien Quiz by comparing children's results on the Alien Quiz with results on others standardised tests of FER for children. At present, the DANVA-CF (Nowicki & Duke, 2013) and NEPSY-II affect recognition (Korkman et al., 2007) are the only well normed instruments of FER in children which have adequate validity (Nowicki & Duke, 2013; Yao et al., 2018) and reliability (Nowicki & Carton, 1993); therefor use of one these instruments is recommended when establishing concurrent validity. The present study investigated validity against more traditional formats, however it is recommended that future research determine concurrent validity through comparison with standardised measures, if possible.

As highlighted earlier no association was found between the Alien Quiz and measures of real-world social functioning. It is recommended that future research attempts to further explore the predictive validity of the task, perhaps using a measure of social adjustment as indicated by Leppänen and Hietanen (2001), or related sociocognitive abilities such as theory of mind or empathy. Comparison between the association that the standardised measure of FER has with real world functioning and the association the Alien Quiz has with real world functioning could provide greater insights into the predictive and ecological validity of the task.

The use of cartoon stimuli may limit the ecological validity of the task. Cartoon expressions of emotion may be more intense than those witnessed in real-world situations. Emotions expressed at higher intensity are more recognisable to children (Cecilione et al., 2017), therefore cartoon stimuli may result in higher accuracy of FER than photographic stimuli, due to the possible exaggeration of emotional expression in cartoon stimuli.

Stimuli for this test was developed with the aim of being culturally fair given the possible confounding effect of culture in tests of FER. Unfortunately, homogeneity of participants meant that cultural fairness could not be assessed in this study. Future research could explore how children from different cultures and ethnicities engage with the Alien Quiz and ascertain the validity of the task with diverse groups of children.

4.7. Research Reflexivity

4.7.1. Professional and Ethical Issues

The process of researching this topic highlighted the relatively limited availability of standardised tests of FER for children; only two measures identified were standardised on children, neither of which are easily accessible. Issues with lack of freely available, easy to administer tests of cognition were raised by Borson et al. (2019) regarding the monetisation of the Montreal Cognitive Assessment. They expressed concerns around lack of accessibility to expensive tests for certain patient

groups, and difficulty paying for accreditation of staff to administer the test. I believe similar concerns are valid in other areas of testing including tests of FER. Both in clinical and research settings, paywall and training difficulties for standardised tests serve as a barrier to their use. It is therefore unsurprising that within research in FER, researchers often created their own novel experimental measures to investigate this construct in children, resulting in issues with comparing and generalising the findings. As was evident in this research, publicly funded research can support the development of new tests of cognition. Nonetheless, without wider support and investment the journey towards standardising these tests is long and arduous.

4.7.2. Personal Reflection

Acknowledgement and consideration on the part of the researcher of the circular relationship between the researcher, social context and history is an important component of reflexivity in research (Flanagan, 1981). In line with a critical realist approach, this research has sought to achieve valid insights into how FER is tested, under the premise that FER is indeed a real construct, while also holding in mind that FER is mediated by language and culture. The adoption of a critical realist approach has undoubtably been shaped by aspects of my own identity such as my family, cultural upbringing, the clinical doctoral training I am completing and the social and political context within which this research has taken place. My identity and experiences have shaped my belief in the importance of assessments to better understand social cognition. A critical realist approach has however, also enabled reflections on the role of language and the social construction of such views. The attribution of Western concepts and approaches to other cultural groups is an ongoing issue within psychology and one which raised personal concerns for me within this research. However, within the current NHS and social care contexts, diagnosis is often used as a barrier to access services, therefore the importance of, developing sensitive and fair approaches to measure difficulties remains.

Although one of the aims of this study was to create a culturally fair test of FER, everyone involved in the development of the task was from a White European background. Reflecting on the process of designing this research, it seems that a

notable oversight was failure to engage with people from other cultural backgrounds on the issue of FER. As researchers we are often advised to look for ‘gaps in the literature’, however it seems that exploration of why this gap is there, and understanding of whether exploration of this is important for those it impacts upon, is also an important question.

The process of completing this research also highlighted biases toward neurotypical presentation in social cognition. The concept that children must be attuned to emotion needs and of others is based upon the prototype of neurotypical children and a Western understanding of how children should behave. This research could be considered as promoting a view of ‘individualised deficits’ whereby some children have impairments which require rectifying. An alternative approach to this area could have been the development of an intervention which supports greater acceptance and support for differences in social behaviour. Such an approach may have enabled exploration of whether changes in school and family responses to difficulties in social cognition would ameliorate challenges experienced in this area among neurodiverse populations.

4.8. Conclusion

The current study explored the utility of a novel measure of FER, designed with children’s developmental stage in mind, and incorporating novel features including a game-like format and culturally neutral stimuli. Findings from this study indicate that the Alien Quiz has potential as a measure of FER for children. Children adopted similar strategies when completing subtests of the tasks, engaging with emotion stimuli in similar ways. This study identified areas in which the task could be improved, including a focus on fewer emotions, and the reduction of other characteristics in the emotion stimuli. Further testing involving a larger sample size, ‘known groups’ impacted by FER difficulties, and bilingual participants, could establish the utility of this task. Findings from this study can be used to guide future research on the Alien Quiz, to develop this task into a standardized tool to detect difficulties in emotion recognition, which could provide benefits for children, clinicians, and researchers.

REFERENCES

Adobe. (2022a). *Adobe Illustrator* (version 26.3.1.). Adobe. https://www.adobe.com/uk/products/illustrator.html

Adobe. (2022b). *Adobe Stock*. Adobe. https://stock.adobe.com/uk

Airdrie, J. N., Langley, K., Thapar, A., & van Goozen, S. H. M. (2018). Facial emotion recognition and eye gaze in Attention-Deficit/Hyperactivity disorder with and without comorbid conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(8), 561–570. https://doi.org/10.1016/j.jaac.2018.04.016

Al-Ababneh, M. M. (2020). Linking ontology, epistemology and research methodology. *Science & Philosophy*, *8*(11). https://papers.ssrn.com/abstract=3708935

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

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association. https://www.psychiatry.org:443/psychiatrists/practice/dsm

Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Villanueva, L. (2016). Social cognition makes an independent contribution to peer relations in children with Specific Language Impairment. *Research in Developmental Disabilities*, *49–50*, 277–290. https://doi.org/10.1016/j.ridd.2015.12.015

Assed, M. M., Khafif, T. C., Belizario, G. O., Fatorelli, R., Rocca, C. C. de A., & de Pádua Serafim, A. (2020). Facial emotion recognition in maltreated children: A systematic review. *Journal of Child and Family Studies*, *29*(5), 1493–1509. https://doi.org/10.1007/s10826-019-01636-w

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

Baron-Cohen, S. (1991). The theory of mind deficit in autism: How specific is it?*. *British Journal of Developmental Psychology*, *9*(2), 301–314. https://doi.org/10.1111/j.2044-835X.1991.tb00879.x

Baron-Cohen, S. (1996). Reading the mind in the face: A cross-cultural and developmental study. *Visual Cognition*, *3*(1), 39–60. https://doi.org/10.1080/713756728

Baron-Cohen, S., Golan, O., Wheelwright, S., & Granader, Y. (2010). Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in Evolutionary Neuroscience*, *2*. https://doi.org/10.3389/fnevo.2010.00109

Barth, J. M., & Bastiani, A. (1997). A longitudinal study of emotion recognition and preschool children's social behavior. *Merrill-Palmer Quarterly*, *43*(1), 107–128.

Bauman, M. L., & Kemper, T. L. (2005). Neuroanatomic observations of the brain in autism: A review and future directions. *International Journal of Developmental Neuroscience*, *23*(2), 183–187. https://doi.org/10.1016/j.ijdevneu.2004.09.006

Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin*, *136*(1), 39–64. https://doi.org/10.1037/a0017768

Beaudoin, C., & Beauchamp, M. H. (2020). Social cognition. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 255–264). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00022-8

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology*, *12*(3), 205–221. https://doi.org/10.1080/09297040500488522

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

Bhavnani, S., Mukherjee, D., Dasgupta, J., Verma, D., Parameshwaran, D., Divan, G., Sharma, K. K., Thiagarajan, T., & Patel, V. (2019). Development, feasibility and acceptability of a gamified cognitive DEvelopmental assessment on an E-Platform (DEEP) in rural Indian pre-schoolers – a pilot study. *Global Health Action*, *12*(1), 1548005. https://doi.org/10.1080/16549716.2018.1548005

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520–532. https://doi.org/10.1016/j.neubiorev.2014.09.021

Bölte, S., Hubl, D., Feineis-Matthews, S., Prvulovic, D., Dierks, T., & Poustka, F. (2006). Facial affect recognition training in autism: Can we animate the fusiform gyrus? *Behavioral Neuroscience*, *120*, 211–216. https://doi.org/10.1037/0735-7044.120.1.211

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*(1), 72–93. https://doi.org/10.1016/0093-934X(85)90029-X

Borson, S., Sehgal, M., & Chodosh, J. (2019). Monetizing the MoCA: What Now? *Journal of the American Geriatrics Society*, *67*(11), 2229–2231. https://doi.org/10.1111/jgs.16158

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of facial emotion expression in major depression: A review. *Australian & New Zealand Journal of Psychiatry*, *44*(8), 681–696. https://doi.org/10.3109/00048674.2010.496359

Bowers, D., Blonder, L. X., & Heilman, K. M. (1998). *Florida affect battery*. Center for Neuropsychological Studies, Department of Neurology.

Brooks, BrianL., Sherman, ElisabethM. S., & Strauss, E. (2010). NEPSY-II: A developmental neuropsychological assessment, second edition. *Child Neuropsychology*, *16*(1), 80–101. https://doi.org/10.1080/09297040903146966

Brotman, M. A., Guyer, A. E., Lawson, E. S., Horsey, S. E., Rich, B. A., Dickstein, D. P., Pine, D. S., & Leibenluft, E. (2008). Facial emotion labeling deficits in children and adolescents at risk for bipolar disorder. *American Journal of Psychiatry*, *165*(3), 385–389. https://doi.org/10.1176/appi.ajp.2007.06122050

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming, second edition* (2nd ed.). Routledge. https://doi.org/10.4324/9780203805534

Camras, L. A., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behavior*, *9*(2), 84–94. https://doi.org/10.1007/BF00987140

Castro, L., Davies, H., Hale, L., Surguladze, S., & Tchanturia, K. (2010). Facial affect recognition in anorexia nervosa: Is obsessionality a missing piece of the puzzle? *The Australian and New Zealand Journal of Psychiatry*, *44*(12), 1118–1125. https://doi.org/10.3109/00048674.2010.524625

Cecilione, J. L., Rappaport, L. M., Verhulst, B., Carney, D. M., Blair, R. J. R., Brotman, M. A., Leibenluft, E., Pine, D. S., Roberson-Nay, R., & Hettema, J. M. (2017). Test–retest reliability of the facial expression labeling task. *Psychological Assessment*, *29*(12), 1537–1542. https://doi.org/10.1037/pas0000439

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, *33*(2), 218–236. https://doi.org/10.1111/bjdp.12075

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement, 3rd ed* (pp. xxviii, 798). Mayfield Publishing Co.

Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: A systematic review. *Research in Developmental Disabilities*, *34*(5), 1505–1520. https://doi.org/10.1016/j.ridd.2013.01.008

Collins. (n.d.). *Collins Thesaurus | Synonyms, Antonyms and Definitions*. https://www.collinsdictionary.com/dictionary/english-thesaurus

Cooper, S., Hobson, C. W., & van Goozen, S. H. (2020). Facial emotion recognition in children with externalising behaviours: A systematic review. *Clinical Child Psychology and Psychiatry*, *25*(4), 1068–1085. https://doi.org/10.1177/1359104520945390

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience & Biobehavioral Reviews*, *84*, 92–99. https://doi.org/10.1016/j.neubiorev.2017.11.014

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830–843. https://doi.org/10.1037/xge0000172

Crotty, M. J. (1998). The foundations of social research: Meaning and perspective in the research process. *The Foundations of Social Research*, 1–256.

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press.

Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Development*, *73*(3), 700–717. https://doi.org/10.1111/1467-8624.00433

de la Torre-Luque, A., Viera-Campos, A., Bilderbeck, A. C., Carreras, M. T., Vivancos, J., Diaz-Caneja, C. M., Aghajani, M., Saris, I. M. J., Raslescu, A., Malik, A., Clark, J., Penninx, B. W. J. H., van der Wee, N., Rossum, I. W., Sommer, B., Marston, H., Dawson, G. R., Kas, M. J., Ayuso-Mateos, J. L., & Arango, C. (2022). Relationships between social withdrawal and facial emotion recognition in neuropsychiatric disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *113*, 110463. https://doi.org/10.1016/j.pnpbp.2021.110463

Dede, B., Delk, L., & White, B. A. (2021). Relationships between facial emotion recognition, internalizing symptoms, and social problems in young children. *Personality and Individual Differences*, *171*, 110448. https://doi.org/10.1016/j.paid.2020.110448

Denham, S., Mitchell-Copeland, J., Srandberg, K., Auerbach, S., & Blair, K. (1997). Parental contributions to preschoolers 'emotional competence: Direct and indirect effects. *Motivation and Emotion*, *21*, 65–86. https://doi.org/10.1023/A:1024426431247

Dyck, M. J. (2012). The ability to understand the experience of other people: Development and validation of the emotion recognition scales. *Australian Psychologist*, *47*(1), 49–57. https://doi.org/10.1111/j.1742-9544.2011.00047.x

Dyck, M. J., Farrugia, C., Shochet, I. M., & Holmes-Brown, M. (2004). Emotion recognition/understanding ability in hearing or vision-impaired children: Do sounds, sights, or words make the difference? *Journal of Child Psychology and Psychiatry*, *45*(4), 789–800. https://doi.org/10.1111/j.1469-7610.2004.00272.x

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., & Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, *72*(4), 1112–1134. https://doi.org/10.1111/1467-8624.00337

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221. https://doi.org/10.1196/annals.1280.010

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Affect*. Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. ISHK.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Emond, A., Ormel, J., Veenstra, R., & Oldehinkel, A. J. (2007). Preschool behavioral and social-cognitive problems as predictors of (pre)adolescent disruptive behavior. *Child Psychiatry and Human Development*, *38*(3), 221–236. https://doi.org/10.1007/s10578-007-0058-5

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

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<375::AID-JHBS2300170308>3.0.CO;2-U

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience*, *33*, 299–324. https://doi.org/10.1146/annurev-neuro-060909-153230

Franco, F., Itakura, S., Pomorska, K., Abramowski, A., Nikaido, K., & Dimitriou, D. (2014). Can children with autism read emotions from the eyes? The Eyes Test revisited. *Research in Developmental Disabilities*, *35*(5), 1015–1026. https://doi.org/10.1016/j.ridd.2014.01.037

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., Baron-Cohen, S., & Golan, O. (2016). Basic and complex emotion recognition in children with autism: Cross-cultural findings. *Molecular Autism*, *7*(1), 52. https://doi.org/10.1186/s13229-016-0113-9

Frith, C. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2033–2039. https://doi.org/10.1098/rstb.2008.0005

Gagliardi, C., Frigerio, E., Burt, D. M., Cazzaniga, I., Perrett, D. I., & Borgatti, R. (2003). Facial expression recognition in Williams syndrome. *Neuropsychologia*, *41*(6), 733–738. https://doi.org/10.1016/S0028-3932(02)00178-1

Gagner, C., Landry-Roy, C., Lainé, F., & Beauchamp, M. H. (2015). Sleep-wake disturbances and fatigue after pediatric traumatic brain injury: A systematic review of the literature. *Journal of Neurotrauma*, *32*(20), 1539–1552. https://doi.org/10.1089/neu.2014.3753

Garretson, H. B., Fein, D., & Waterhouse, L. (1990). Sustained attention in children with autism. *Journal of Autism and Developmental Disorders*, *20*(1), 101–114. https://doi.org/10.1007/BF02206860

Garrigan, B., Adlam, A. L. R., & Langdon, P. E. (2016). The neural correlates of moral decision-making: A systematic review and meta-analysis of moral

evaluations and response decision judgements. *Brain and Cognition*, *108*, 88–97. https://doi.org/10.1016/j.bandc.2016.07.007

Gioia, G., & Isquith, P. (2011). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, *25*(1–2), 135–158.

Golan, O., Sinai-Gavrilov, Y., & Baron-Cohen, S. (2015). The Cambridge Mindreading Face-Voice Battery for Children (CAM-C): Complex emotion recognition in children with and without autism spectrum conditions. *Molecular Autism*, *6*(1), 22. https://doi.org/10.1186/s13229-015-0018-z

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

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, *7*(11), 125–130. https://doi.org/10.1007/s007870050057

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *American Psychologist*, *69*(1), 66–78. https://doi.org/10.1037/a0034857

Hair, J. F. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson Education. http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&doc_number=016464401&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

Hawkins, G. E., Rae, B., Nesbitt, K. V., & Brown, S. D. (2013). Gamelike features might not improve data. *Behavior Research Methods*, *45*(2), 301–318. https://doi.org/10.3758/s13428-012-0264-3

Henry, J. D., von Hippel, W., Molenberghs, P., Lee, T., & Sachdev, P. S. (2016). Clinical assessment of social cognitive function in neurological disorders. *Nature Reviews Neurology*, *12*(1), 28–39. https://doi.org/10.1038/nrneurol.2015.229

Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., Santosh, P., & Phillips, M. (2008). Impact of familiarity upon children's

developing facial expression recognition. *Journal of Child Psychology and Psychiatry*, *49*(2), 201–210. https://doi.org/10.1111/j.1469-7610.2007.01835.x

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, *47*(11), 1098–1106. https://doi.org/10.1111/j.1469-7610.2006.01652.x

Howlin, P., Baron-Cohen, S., & Hadwin, J. A. (1999). *Teaching children with autism to mind-read: A practical guide for teachers and parents*. John Wiley & Sons.

IBM Corp. (2020). *IBM SPSS statistics for macintosh, version 27.0.* (27.0). IBM Corp.

Joseph, R. M., & Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 529–542. https://doi.org/10.1111/1469-7610.00142

Kazanas, S. A., McLean, J. S., & Altarriba, J. (2019). Emotion and Emotion Concepts. In *The Handbook of the Neuroscience of Multilingualism* (pp. 313–334). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119387725.ch15

Kerns, K. A. (2000). The CyberCruiser: An investigation of development of prospective memory in children. *Journal of the International Neuropsychological Society*, *6*(1), 62–70. https://doi.org/10.1017/S1355617700611074

Kessels, R. P. C., Montagne, B., Hendriks, A. W., Perrett, D. I., & de Haan, E. H. F. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology*, *8*(1), 75–93. https://doi.org/10.1111/jnp.12009

Kessler, H., Bayerl, P., Deighton, R. M., & Traue, H. C. (2002). Facially expressed emotion labeling (FEEL): PC-gestützter test zur Emotionserkennung. *Verhaltenstherapie Und Verhaltensmedizin, 23*(3), 297–306. https://doi.org/10.1037/t77311-000

Khawar, R., Malik, F., Maqsood, S., Yasmin, T., & Habib, S. (2014). Age and gender differences in emotion recognition ability and intellectual functioning. *Journal of Behavioral Sciences*.

Kilford, E. J., Garrett, E., & Blakemore, S.-J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience &*

Biobehavioral Reviews, *70*, 106–120. https://doi.org/10.1016/j.neubiorev.2016.08.016

Kobiella, A., Grossmann, T., Reid, V. M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition and Emotion*, *22*(1), 134–146. https://doi.org/10.1080/02699930701394256

Koenig, M. A., & Harris, P. L. (2005). The role of social cognition in early trust. *Trends in Cognitive Sciences*, *9*(10), 457–459. https://doi.org/10.1016/j.tics.2005.08.006

Kolb, B., Wilson, B., & Taylor, L. (1992). Developmental changes in the recognition and comprehension of facial expression: Implications for frontal lobe function. *Brain and Cognition*, *20*(1), 74–84. https://doi.org/10.1016/0278-2626(92)90062-Q

Korkman, M., Kirk, U., & Kemp, S. (2012). *NEPSY - Second Edition*. Pearson. https://doi.org/10.1037/t15125-000

Korkmaz, B. (2011). Theory of mind and neurodevelopmental disorders of childhood. *Pediatric Research*, *69*(8), 101–108. https://doi.org/10.1203/PDR.0b013e318212c177

Kristen, S., Sodian, B., Thoermer, C., & Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. *Developmental Psychology*, *47*(5), 1207–1219. https://doi.org/10.1037/a0024808

Kujawa, A., Dougherty, L., Durbin, C. E., Laptook, R., Torpey, D., & Klein, D. N. (2014). Emotion recognition in preschool children: Associations with maternal depression and early parenting. *Development and Psychopathology*, *26*(1), 159–170. https://doi.org/10.1017/S0954579413000928

Kwan, K. J., Oh, S. W., & Kim, C. T. (2011). *Korean version Wechsler intelligence scale for children-IV.* HakJiSa.

Langenecker, S. A., Bieliauskas, L. A., Rapport, L. J., Zubieta, J.-K., Wilde, E. A., & Berent, S. (2005). Face emotion perception and executive functioning deficits in depression. *Journal of Clinical and Experimental Neuropsychology*, *27*(3), 320–333. https://doi.org/10.1080/13803390490490515720

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*, *6*. https://doi.org/doi.org/10.3389/fpsyg.2015.00761

Lázaro, E., Amayra, I., López-Paz, J. F., Martínez, O., Pérez, M., Berrocoso, S., García, M., Jometón, A., Al-Rashaida, M., Oliva, M., Parada, P., & Hoffmann, H. (2019). Instrument for assessing the ability to identify emotional facial expressions in healthy children and in children with ADHD: The FEEL Test. *Journal of Attention Disorders*, *23*(6), 563–569. https://doi.org/10.1177/1087054716682335

Leime, J. L., Neto, J. R., Alves, S. M., & Torro-Alves, N. (2013). Recognition of facial expressions in children, young adults and elderly people. *Estudos de Psicologia*, *30*(2), 161–167. https://doi.org/10.1590/S0103-166X2013000200002

Leppänen, J. M. (2011). Neural and Developmental Bases of the Ability to Recognize Social Signals of Emotions. *Emotion Review*, *3*(2), 179–188. https://doi.org/10.1177/1754073910387942

Leppänen, J. M., & Hietanen, J. K. (2001). Emotion recognition and social adjustment in school–aged girls and boys. *Scandinavian Journal of Psychology*, *42*(5), 429–435. https://doi.org/10.1111/1467-9450.00255

Lewis, Z. H., Swartz, M. C., & Lyons, E. J. (2016). What's the point?: A review of reward systems implemented in gamification interventions. *Games for Health Journal*, *5*(2), 93–99. https://doi.org/10.1089/g4h.2015.0078

Lopatina, O. L., Komleva, Y. K., Gorina, Y. V., Higashida, H., & Salmina, A. B. (2018). Neurobiological Aspects of Face Recognition: The Role of Oxytocin. *Frontiers in Behavioral Neuroscience*, *12*, 195. https://doi.org/10.3389/fnbeh.2018.00195

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2), e5888. https://doi.org/10.2196/games.5888

MacDonald, P. M., Kirkpatrick, S. W., & Sullivan, L. A. (1996). Schematic drawings of facial expressions for emotion recognition and interpretation by preschool-

aged children. *Genetic, Social, and General Psychology Monographs*, *122*(4), 373–388.

Maire, J., Galera, C., Roux, S., Bioulac, S., Bouvard, M., & Michel, G. (2019). Facial emotion recognition in children with or without Attention Deficit/Hyperactivity Disorder: Impact of comorbidity. *L'Encéphale*, *45*(2), 114–120. https://doi.org/10.1016/j.encep.2018.01.006

Martinovic, D., Burgess, G. H., Pomerleau, C. M., & Marin, C. (2015). Comparison of children's gaming scores to NEPSY-II scores: Validation of computer games as cognitive tools. *Computers in Human Behavior*, *49*, 487–498. https://doi.org/10.1016/j.chb.2015.03.039

McClure, E. B., Treland, J. E., Snow, J., Schmajuk, M., Dickstein, D. P., Towbin, K. E., Charney, D. S., Pine, D. S., & Leibenluft, E. (2005). Deficits in social cognition and response flexibility in pediatric bipolar disorder. *American Journal of Psychiatry*, *162*(9), 1644–1651. https://doi.org/10.1176/appi.ajp.162.9.1644

McCormick, E. M., van Hoorn, J., Cohen, J. R., & Telzer, E. H. (2018). Functional connectivity in the social brain across childhood and adolescence. *Social Cognitive and Affective Neuroscience*, *13*(8), 819–830. https://doi.org/10.1093/scan/nsy064

McKown, C., Russo-Ponsaran, N. M., Johnson, J. K., Russo, J., & Allen, A. (2016). Web-based assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, *34*(4), 322–338. https://doi.org/10.1177/0734282915604564

McPherson, J., & Burns, N. R. (2007). Gs Invaders: Assessing a computer game-like test of processing speed. *Behavior Research Methods*, *39*(4), 876–883. https://doi.org/10.3758/BF03192982

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

Missaghi-Lakshman, M., & Whissell, C. (1991). Children's understanding of facial expression of emotion: II. Drawing of emotion-faces. *Perceptual and Motor Skills*, *72*(3_suppl), 1228–1230. https://doi.org/10.2466/pms.1991.72.3c.1228

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Social Development*, *19*(1), 71–92. https://doi.org/10.1111/j.1467-9507.2008.00527.x

Morrison, K. E., DeBrabander, K. M., Jones, D. R., Ackerman, R. A., & Sasson, N. J. (2020). Social cognition, social skill, and social motivation minimally predict social interaction outcomes for autistic and non-autistic adults. *Frontiers in Psychology*, *11*, 3282. https://doi.org/10.3389/fpsyg.2020.591100

Mukherjee, D., Bhavnani, S., Swaminathan, A., Verma, D., Parameshwaran, D., Divan, G., Dasgupta, J., Sharma, K., Thiagarajan, T. C., & Patel, V. (2020). Proof of concept of a gamified DEevelopmental assessment on an E-Platform (DEEP) tool to measure cognitive development in rural Indian preschool children. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01202

Nelson, E. E., & Guyer, A. E. (2011). The development of the ventral prefrontal cortex and social flexibility. *Developmental Cognitive Neuroscience*, *1*(3), 233–245. https://doi.org/10.1016/j.dcn.2011.01.002

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

Nguyen, C. T. (2017). Philosophy of games. *Wiley Online Library*, *12*(8), e12426. https://doi.org/doi.org/10.1111/phc3.12426

Nowicki, S., Bliwise, N., & Joinson, C. (2019). The association of children's locus of control orientation and emotion recognition abilities at 8 years of age and teachers' ratings of their personal and social difficulties at 10 years. *Journal of Nonverbal Behavior*, *43*(3), 381–396. https://doi.org/10.1007/s10919-019-00304-3

Nowicki, S., & Carton, J. (1993). The measurement of emotional intensity from facial expressions. *The Journal of Social Psychology*, *133*(5), 749–750. https://doi.org/10.1080/00224545.1993.9713934

Nowicki, S., & Duke, M. (2001). Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA). *Interpersonal Sensitivity: Theory and Measurement*, 183–198. https://doi.org/10.4324/9781410600424-19

Nowicki, S., & Duke, M. P. (2013). *Manual for the receptive tests of the diagnostic analysis of nonverbal accuracy 2 (DANVA2)*.

Obradović, J., Sulik, M. J., Finch, J. E., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology*, *55*, 4–13. https://doi.org/10.1016/j.appdev.2017.03.003

Operto, F. F., Pastorino, G. M. G., Stellato, M., Morcaldi, L., Vetri, L., Carotenuto, M., Viggiano, A., & Coppola, G. (2020). Facial Emotion Recognition in Children and Adolescents with Specific Learning Disorder. *Brain Sciences*, *10*(8), Article 8. https://doi.org/10.3390/brainsci10080473

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pahnke, R., Mau-Moeller, A., Hamm, A. O., & Lischke, A. (2020). Reading the mind in the eyes of children test (RME-C-T): Development and validation of a complex emotion recognition test. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00376

Paine, A. L., van Goozen, S. H. M., Burley, D. T., Anthony, R., & Shelton, K. H. (2021). Facial emotion recognition in adopted children. *European Child & Adolescent Psychiatry*. https://doi.org/10.1007/s00787-021-01829-z

Paiva-Silva, A. I. de, Pontes, M. K., Aguiar, J. S. R., & de Souza, W. C. (2016). How do we evaluate facial emotion recognition? *Psychology & Neuroscience*, *9*(2), 153–175. https://doi.org/10.1037/pne0000047

Pavitt, A. (2017). *Using a game-like procedure as a new test of problem solving and concept formation in children* [Pro_doc, University of East London]. https://doi.org/10.15123/PUB.6736

Peters, A. T., Ren, X., Bessette, K. L., George, N., Kling, L. R., Thies, B., West, A. E., Langenecker, S. A., & Pandey, G. N. (2021). Inflammation, depressive symptoms, and emotion perception in adolescence. *Journal of Affective Disorders*, *295*, 717–723. https://doi.org/10.1016/j.jad.2021.08.126

Piedmont, R. L. (2014). Inter-item Correlations. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493

Pitchford, N. J., & Mullen, K. T. (2001). Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology*, *80*(3), 289–314. https://doi.org/10.1006/jecp.2001.2634

Pitchford, N. J., & Outhwaite, L. A. (2016). Can touch screen tablets be used to assess cognitive and motor skills in early years primary school children? A cross-cultural study. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01666

Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology*, *36*, 679–688. https://doi.org/10.1037/0012-1649.36.5.679

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152. https://doi.org/10.1080/17405620344000022

Rapport, L. J., Friedman, S. L., Tzelepis, A., & Van Voorhis, A. (2002). Experienced emotion and affect recognition in adult attention-deficit hyperactivity disorder. *Neuropsychology*, *16*(1), 102–110. https://doi.org/10.1037/0894-4105.16.1.102

Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, *18*(6), 926–939. https://doi.org/10.1111/desc.12281

Rodgers, J. D., Lopata, C., Booth, A. J., Thomeer, M. L., Donnelly, J. P., Rajnisz, C. J., Wood, J. T., Lodi-Smith, J., & Kozlowski, K. F. (2021). Psychometric properties of the Cambridge-Mindreading Face-Voice Battery for Children in children with ASD. *Autism Research*, *14*(9), 1965–1974. https://doi.org/10.1002/aur.2546

Rosetti, M. F., Gómez-Tello, M. F., Victoria, G., & Apiquian, R. (2017). A video game for the neuropsychological screening of children. *Entertainment Computing*, *20*, 1–9. https://doi.org/10.1016/j.entcom.2017.02.002

Rueda, P., Cabello, R., & Fernández-Berrocal, P. (2013). Preliminary validation of spanish 'Eyes test-child version': Validación preliminar del 'Test de los Ojos para niños' en castellano. *Ansiedad y Estrés*, *19*(2/3), 173–184.

Salen, K., & Zimmerman, E. (2003). *Rules of Play*. The MIT Press. https://mitpress.mit.edu/9780262240451/rules-of-play/

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson Education.

Scherer, K., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition index. *Journal of Nonverbal Behavior*, *35*(4), 305–326. https://doi.org/10.1007/s10919-011-0115-4

Schmidt, A. T., Hanten, G. R., Li, X., Orsten, K. D., & Levin, H. S. (2010). Emotion recognition following pediatric traumatic brain injury: Longitudinal analysis of emotional prosody and facial emotion recognition. *Neuropsychologia*, *48*(10), 2869–2877. https://doi.org/10.1016/j.neuropsychologia.2010.05.029

Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, *23*(2), 125–141. https://doi.org/10.1016/j.ijdevneu.2004.12.012

Schworer, E. K., Hoffman, E. K., & Esbensen, A. J. (2021). Psychometric evaluation of social cognition and behavior measures in children and adolescents with Down Syndrome. *Brain Sciences*, *11*(7), Article 7. https://doi.org/10.3390/brainsci11070836

Shanok, N. A., Jones, N. A., & Lucas, N. N. (2019). The nature of facial emotion recognition impairments in children on the autism spectrum. *Child Psychiatry & Human Development*, *50*(4), 661–667. https://doi.org/10.1007/s10578-019-00870-z

Silva, T. B. F., Osório, F. L., & Loureiro, S. R. (2015). SDQ: Discriminative validity and diagnostic potential. *Frontiers in Psychology*, *6*, 811. https://doi.org/10.3389/fpsyg.2015.00811

Simcock, G., McLoughlin, L. T., De Regt, T., Broadhouse, K. M., Beaudequin, D., Lagopoulos, J., & Hermens, D. F. (2020). Associations between facial emotion recognition and mental health in early adolescence. *International Journal of Environmental Research and Public Health*, *17*(1), Article 1. https://doi.org/10.3390/ijerph17010330

Simonian, S. J., Beidel, D. C., Turner, S. M., Berkes, J. L., & Long, J. H. (2001). Recognition of facial affect by children and adolescents diagnosed with social phobia. *Child Psychiatry and Human Development*, *32*(2), 137–145. https://doi.org/10.1023/A:1012298707253

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, *21*(2), 345–354. https://doi.org/10.1016/S0163-6383(98)90011-X

Snodgrass, C., & Knott, F. (2006). Theory of mind in children with traumatic brain injury. *Brain Injury*, *20*(8), 825–833. https://doi.org/10.1080/02699050600832585

Somerville, L. H., Fani, N., & McClure-Tone, E. B. (2011). Behavioral and neural representation of emotional facial expressions across the lifespan. *Developmental Neuropsychology*, *36*(4), 408–428. https://doi.org/10.1080/87565641.2010.549865

Song, H., Yi, D.-J., & Park, H.-J. (2020). Validation of a mobile game-based assessment of cognitive control among children and adolescents. *PLoS ONE*, *15*(3), 1–18. https://doi.org/10.1371/journal.pone.0230498

Sowell, E. R., Trauner, D. A., Gamst, A., & Jernigan, T. L. (2002). Development of cortical and subcortical brain structures in childhood and adolescence: A structural MRI study. *Developmental Medicine and Child Neurology*, *44*(1), 4–16. https://doi.org/10.1017/S0012162201001591

Staff, A. I., Luman, M., van der Oord, S., Bergwerff, C. E., van den Hoofdakker, B. J., & Oosterlaan, J. (2022). Facial emotion recognition impairment predicts social and emotional problems in children with (subthreshold) ADHD. *European Child & Adolescent Psychiatry*, *31*(5), 715–727. https://doi.org/10.1007/s00787-020-01709-y

Staphorst, M. S., Benninga, M. A., Bisschoff, M., Bon, I., Busschbach, J. J. V., Diederen, K., Goudoever, J. B. van, Haarman, E. G., Hunfeld, J. A. M., Jaddoe, V. V. W., Jong, K. J. M. de, Jongste, J. C. de, Kindermann, A., Königs, M., Oosterlaan, J., Passchier, J., Pijnenburg, M. W., Reneman, L., Ridder, L. de, … Vathorst, S. van de. (2017). The child's perspective on discomfort during medical research procedures: A descriptive study. *BMJ Open*, *7*(7), e016077. https://doi.org/10.1136/bmjopen-2017-016077

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, *13*(3), 254–274. https://doi.org/10.1007/s10567-010-0071-2

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Ournal of Experimental Psychology*, *36*, 684–700.

Swaab-barneveld, H., de SONNEVILLE, L., Cohen-kettenis, P., Gielen, A., Buitelaar, J., & van ENGELAND, H. (2000). Visual sustained attention in a child psychiatric population. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(5), 651–659. https://doi.org/10.1097/00004583-200005000-00020

Szanto, K., Dombrovski, A. Y., Sahakian, B. J., Mulsant, B. H., Houck, P. R., Reynolds, C. F., & Clark, L. (2012). Social emotion recognition, social functioning, and attempted suicide in late-life depression. *The American Journal of Geriatric Psychiatry*, *20*(3), 257–265. https://doi.org/10.1097/JGP.0b013e31820eea0c

Taamallah, A., Halayem, S., Rajhi, O., Ghazzai, M., Moussa, M., Touati, M., Ayadi, H. B. Y., Ouanes, S., Abbes, Z. S., Hajri, M., Jelili, S., Fakhfakh, R., & Bouden, A. (2021). Validation of the Tunisian test for facial emotions recognition: Study in children from 7 to 12 years old. *Frontiers in Psychology*, *12*. https://doi.org/doi.org/10.3389/fpsyg.2021.643749

Tenorio Delgado, M., Arango Uribe, P., Aparicio Alonso, A., & Rosas Díaz, R. (2016). TENI: A comprehensive battery for cognitive assessment based on games and technology. *Child Neuropsychology*, *22*(3), 276–291. https://doi.org/10.1080/09297049.2014.977241

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007a). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, *10*(5), 547–558. https://doi.org/10.1111/j.1467-7687.2007.00614.x

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007a). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury*, *21*(6), 623–629. https://doi.org/10.1080/02699050701426865

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007b). Reading emotions after child brain injury: A comparison between children with brain injury and non-injured controls. *Brain Injury*, *21*(7), 731–739. https://doi.org/10.1080/02699050701426899

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Trochim, W. M. K. (2001). *Research methods knowledge base* (2nd ed). Atomic Dog Publisher.

Vivanti, G., Fanning, P. A. J., Hocking, D. R., Sievers, S., & Dissanayake, C. (2017). Social attention, joint attention and sustained attention in autism spectrum disorder and Williams Syndrome: Convergences and sivergences. *Journal of Autism and Developmental Disorders*, *47*(6), 1866–1877. https://doi.org/10.1007/s10803-017-3106-4

Wagner, M. F., Milner, J. S., McCarthy, R. J., Crouch, J. L., McCanne, T. R., & Skowronski, J. J. (2015). Facial emotion recognition accuracy and child physical abuse: An experiment and a meta-analysis. *Psychology of Violence*, *5*, 154–162. https://doi.org/10.1037/a0036014

Waters, A. M., Neumann, D. L., Henry, J., Craske, M. G., & Ornitz, E. M. (2008). Baseline and affective startle modulation by angry and neutral faces in 4–8-year-old anxious and non-anxious children. *Biological Psychology*, *78*(1), 10–19. https://doi.org/10.1016/j.biopsycho.2007.12.005

Weber, R. P. (1990). *Basic content analysis*. SAGE.

Wells, A. E., Hunnikin, L. M., Ash, D. P., & van Goozen, S. H. M. (2020). Children with behavioural problems misinterpret the emotions and intentions of others. *Journal of Abnormal Child Psychology*, *48*(2), 213–221. https://doi.org/10.1007/s10802-019-00594-7

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664. https://doi.org/10.1016/s0896-6273(03)00679-2

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*, 114–128. https://doi.org/10.1037/0012-1649.39.1.114

Williams, L. M., Hermens, D. F., Palmer, D., Kohn, M., Clarke, S., Keage, H., Clark, C. R., & Gordon, E. (2008). Misinterpreting emotional expressions in Attention-Deficit/Hyperactivity Disorder: Evidence for a neural marker and stimulant effects. *Biological Psychiatry*, *63*(10), 917–926. https://doi.org/10.1016/j.biopsych.2007.11.022

Willig, C. (2008). *Introducing Qualitative Research in Psychology*. Open University Press.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283. https://doi.org/10.1038/nn816

Woods, S., Wolke, D., Nowicki, S., & Hall, L. (2009). Emotion recognition abilities and empathy of victims of bullying. *Child Abuse & Neglect*, *33*(5), Article 5. https://doi.org/10.1016/j.chiabu.2008.11.002

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Yao, S.-Y., Bull, R., Khng, K. H., & Rahim, A. (2018). Psychometric properties of the NEPSY-II affect recognition subtest in a preschool sample: A Rasch modeling approach. *The Clinical Neuropsychologist*, *32*(1), 63–80. https://doi.org/10.1080/13854046.2017.1343865

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST)*. Thames Valley Test Company.

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology*, *17*(2), 251–257. https://doi.org/10.1016/j.conb.2007.03.009

Zonnevijlle-Bendek, M. J. S., van Goozen, S. H. M., Cohen-Kettenis, P. T., van Elburg, A., & van Engeland, H. (2002). Do adolescent anorexia nervosa patients have deficits in emotional functioning? *European Child & Adolescent Psychiatry*, *11*(1), 38–42. https://doi.org/10.1007/s007870200006

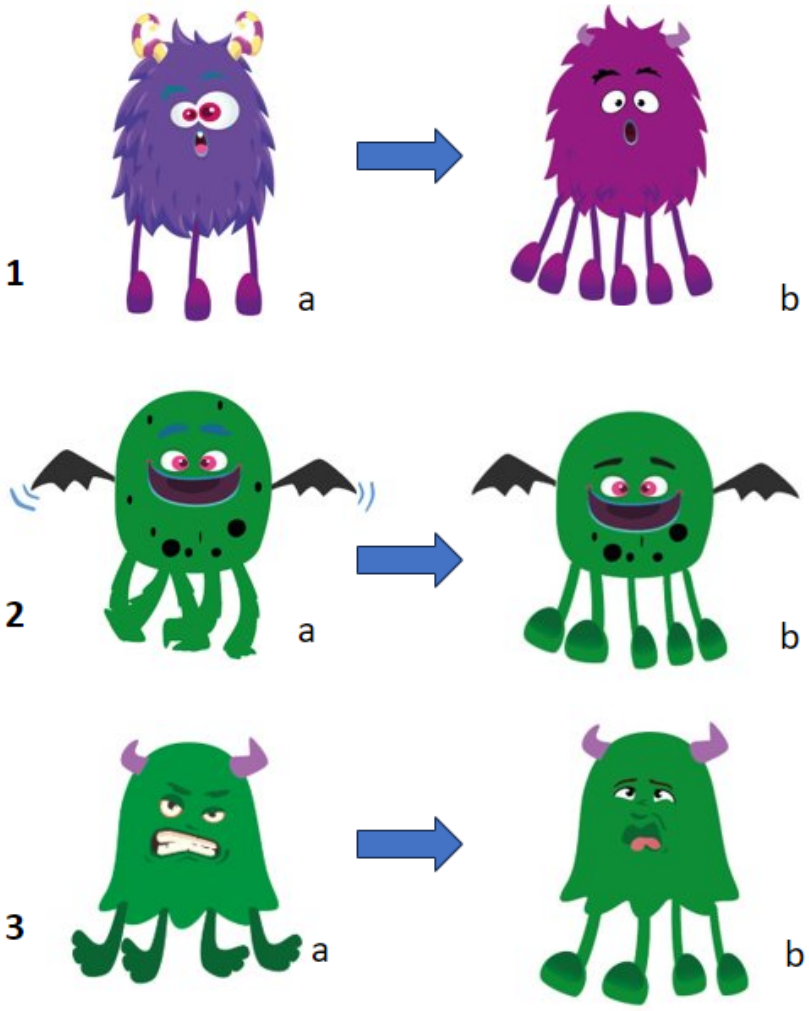
APPENDICES

APPENDIX A: Stimuli Development Scheme

frequency	characteristic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 in 3	Horns	y	n	n	y	n	n	y	n	n	y	n	n	y	n	n	y	n	n	y	n	n	y	n	n
1 in 3	Tail	n	y	n	n	y	n	n	n	n	y	y	n	n	n	y	n	n	Y	n	y	n	n	y	n
1 in 2	Tongue	n	n	y	y	n	n	y	n	y	n	y	y	y	n	y	n	n	y	n	n	n	y	n	y
1 in 3	Skin	F	Sp	Sm	Sm	Sp	F	F	Sp	Sm	Sm	Sp	F	F	Sp	Sm	Sm	Sp	F	F	Sm	Sp	F	Sp	Sm
1 in 7	Colour	O	G	R	G	R	G	B	Y	P	G	Y	R	G	P	P	B	G	G	P	O	Y	G	O	B
1 in 7	Wings	n	n	Y	n	n	n	n	n	n	n	n	n	N	y	n	n	y	n	n	n	n	n	n	n
1 in 7	Legs	3	2	0	4	5	1	3	0	6	4	5	3	1	2	6	4	5	1	6	2	3	4	0	1
1 in 7	Emotion	Su	N	F	D	J	Sa	Su	A	N	J	D	Sa	Su	A	F	Sa	J	N	Su	A	F	D	J	D

Table Key
Horns, Tail, Tongue, Wings: present in Alien yes (y) or no (n)
Skin: Fur=F, Spot=Sp, Smooth=Sm
Colour: O = orange, G = grey, R = red, G = green, B = blue, Y = yellow, P = purple
Legs: number of legs alien have
Emotion: Su = surprise, N = neutral, F = fear, D = disgust, J = joy, Sa = sadness, A = anger

APPENDIX B: Example Alterations to Emotion Stimuli During Piloting



1: Feedback during piloting on both adults and children suggested that the original image (a) could be interpreted as 'confused' due to the difference in height of eyebrows and difference in eye size. Image b represents the final 'surprise' emotion stimuli.

2: Movement lines for example item 2 were removed as pilot feedback suggested that this indicated 'excitement'**.**

3. A protruding tongue and wrinkled nose bridge were introduced for the 'disgust' emotion stimuli following piloting.

APPENDIX C: Child Information Sheet

INVITATION TO TAKE PART IN RESEARCH

University of East London

Using a game-like task as an assessment of emotion recognition in children

You are being invited to take part in some research. You do not have to take part if you do not want to. Before you agree it is important that you understand what it would involve. Please take the time to read this with your parent or caregiver.

Who am I?

My name is Méabh Foley, I am a Trainee Clinical Psychologist.
I am doing some research as part of my studies at University.

What is the purpose of the research?

Sometimes our brains work differently from one another, and it can be helpful to measure how our brain is working. We need good measures to assess this so we can learn more about our brains and how we think about others.

I have made a game called The Alien Quiz. I want to know if young people your age enjoy this game, and whether it can tell me anything about the way your brain works.

Why you?

I want to see whether the new game works properly for children like yourself. That is why I am looking for children just like you to take part.

You and the person who looks after you can decide whether you would like to take part. It is completely up to you!

What will you be asked to do?

You will meet with me and I will ask you to play the game with me, the game will use pictures of aliens that you can ask me questions about. I will then ask you question about how much you enjoyed the game. I will also ask your teacher some questions about you, which will help us to assess how good the new game is.

What happens to your information?

Any information you tell me will be anonymised, which means rather than recording your name I will give you a number, so no one will know it is your information.

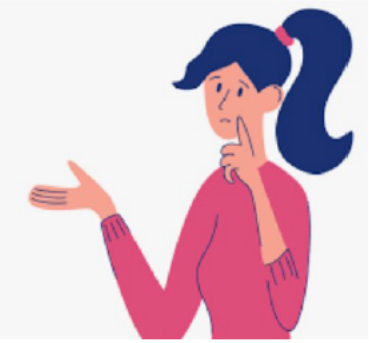
The information will be stored in an electronic cloud with a password only I will know. I will look at the information with my supervisor, who I work with. The information will then be put into writing for other psychologists to read.

Do I have to take part?
No. It is up to you to decide if you want to take part.
If you do not want to, that is fine.

What will happen to the results of the research?
I will look at the information with my supervisor, who I work with. The information will then be put into writing for other psychologists to read.

What if you change your mind?

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



Contact Details

If you have any questions you can ask the person who looks after you to email me. My email address is m.foley2001@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)

APPENDIX D: Child Consent Form

The University of East London

Using a game-like task as an assessment of emotion recognition in children

Consent to participate in a research study

Please carefully read each statement and tick the box if you agree

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

I would like to receive a summary of the research once the study has finished and will ask my parent or caregiver to send contact details for this to be sent to. ☐

I agree to take part in the study. ☐

Your Name (BLOCK CAPITALS)

……

Your Signature

……

Researcher's Name (BLOCK CAPITALS)

……

Researcher's Signature

……

Date: ……

APPENDIX E: Child Debrief Sheet

University of East London

PARTICIPANT DEBRIEF SHEET

Using a game-like task as an assessment of emotion recognition in children

Thank you for taking part in my study!

This study was looking at whether young people your age enjoy The Alien Quiz, and whether it can tell me 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?

I will look at the information with my supervisor, who I work with. The information will then be put into writing for other psychologists to read.

Any information you told me will be anonymised, which means rather than recording your name I 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 it is possible that you may have found something hard or distressing or uncomfortable in some way. If you have been affected in any of those ways, you may find Childline helpful for guidance and support. Childline can be contacted on 0800 1111.

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 me. My email address is m.foley2001@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)

Thank you again for taking part in my study!

APPENDIX F: Record Form

Aliens Quiz Record Form

Demographic Details

ID Number:
DoB:
Age:
Sex/GI:
Nationality/Ethnicity:
Primary language:
Other language(s):
Sensory or motor needs:
Have you tried a game like this before?
Test by:
Test date:
Test location:
Notes:

Record & Observations Sheet

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

Other questions or queries:

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

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

Other questions or queries:

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

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

Other questions or queries:

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

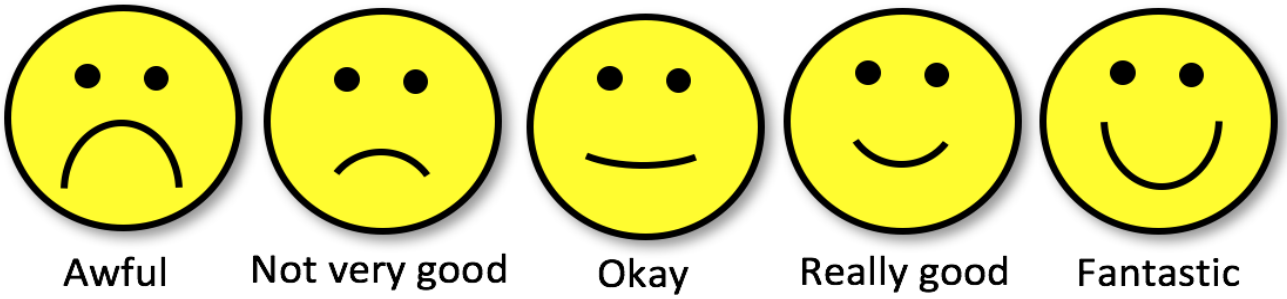
Trial D Target Alien:
Game questions asked: (questions that elicit a yes or no response)
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Other questions or queries:

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

How enjoyable was the task today?

Point to the face that shows how you felt:



Verbal Instructions: Ask how enjoyable Guess the Alien and Emotion Matching tasks were. Mark response on sheet.

1. Affect Naming

SAY: *"To start, have a look at this alien* [point to number 5]. *Tell me, how is this alien feeling?"*
If child cannot answer or says they do not understand give further help as necessary; for example, SAY *"Looking at the face, what do you think this alien is feeling today? What emotion is this alien showing?"*
Record response verbatim.
Continue as above for each of the five remaining emotions

Number	Response	Score
5		0 1
20		0 1
22		0 1
16		0 1
3		0 1
1		0 1

Affect naming: score 1 point for each correct response or close synonym

2. Affect Recognition

(a) SAY: *"Great. Now, look again at this alien* [point to number 5]. *Tell me, is the alien feeling,* ***angry****,* ***afraid****,* ***happy*** *or* ***surprised****?"*
Repeat instruction if necessary.
Record response verbatim and note each correct response.
Continue as above for each of the five remaining emotions, phrasing as follows (each emotion appears 4 times overall in these questions):
(b) *"Look again at this alien* [point to number 20]. *Tell me, is the alien feeling,* ***sad****,* ***angry****,* ***happy*** *or* ***disgusted****?"*
(c) *"Look again at this alien* [point to number 22]. *Tell me, is the alien feeling,* ***surprised****,* ***sad****,* ***afraid*** *or* ***disgusted****?"*
(d) *"Look again at this alien* [point to number 16]. *Tell me, is the alien feeling,* ***sad****,* ***happy****,* ***afraid****, or* ***surprised****?"*
(e) *"Look again at this alien* [point to number 3]. *Tell me, is the alien feeling,* ***angry****,* ***afraid****,* ***disgusted****, or* ***happy****?"*
(f) *"Look again at this alien* [point to number 1]. *Tell me, is the alien feeling,* ***disgusted****,* ***angry****,* ***sad*** *or* ***surprised****?"*

Number	Response (circle)					Score
5	angry	afraid	**happy**	Surprised	DK	0 1
20	sad	**angry**	happy	disgusted	DK	0 1
22	surprised	sad	afraid	**disgusted**	DK	0 1
16	**sad**	happy	afraid	surprised	DK	0 1
3	angry	**afraid**	disgusted	happy	DK	0 1
1	disgusted	angry	sad	**surprised**	DK	0 1
Total						

Affect recognition: score 1 point for each correct response

3. Affect Matching

SAY: *"Great. Now, look again at this alien* [point to number 5] *then look at the other aliens on the card. Point out to me as many aliens as you can who are feeling the same way as this alien. Which other aliens are showing the same emotion? Have a look and point them out to me."*

Record aliens identified, either by number or by pointing.

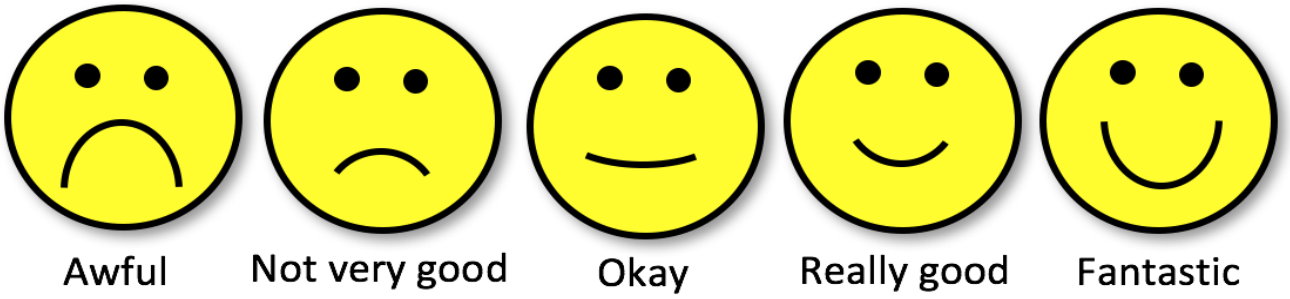
Continue as above for each of the five remaining emotions.

Alien	Response	Correct response (circle)			Score
Alien 5 (happy)		10	17	23	0 1 2 3
Alien 20 (angry)		8	14		0 1 2
Alien 22 (disgusted)		4	11	24	0 1 2 3
Alien 16 (sad)		6	12		0 1 2
Alien 3 (afraid)		15	21		0 1 2
Alien 1 (surprised)		7	13	19	0 1 2 3
Total					

Affect matching: score 1 point for each alien correctly identified.

How enjoyable was the task today?

Point to the face that shows how you felt:



Verbal Instructions: Ask how enjoyable Guess the Alien and Emotion Matching tasks were. Mark response on sheet.

APPENDIX G: Instructions Sheet Guess the Alien

Script for Guess the Alien instructions.

SAY: *"Have you ever played the game guess who before? No, that's OK I will teach you how to play. I want you to try and guess which one of these aliens [point towards aliens] I have in my hand. To figure out which alien I have you can ask me questions about the aliens. You can only ask me questions that I can answer yes or no to. Don't ask any questions about the numbers. You can flip the card over, like this* [demonstrate], *of the aliens that you know aren't right after you ask the question. To win the game you have to try and guess which alien I have in less than 10 questions."*

If child cannot answer or says they do not understand give further help as necessary, for example, SAY *"Can you think of a question about what the alien looks like?"*

If child asks open ended question SAY *"I can only answer yes or no would you like to try another question that I can answer yes or no to"*

If child still does not understand SAY *"/an example question would be does the alien have fur, if the answer is yes you flip over the cards of the aliens who do not have fur"*

APPENDIX H: Peer Problem and Prosocial Subscale of SDQ

For teacher to complete
Please return to Meabh Foley, Trainee Clinical Psychologist

Child's Name .....
Male/Female

Strengths and Difficulties Questionnaire

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of the child's behaviour over the last six months or this school year.

	Not True	Somewhat True	Certainly True
Rather solitary, tends to play alone			
Has at least one good friend			
Generally liked by other children			
Picked on or bullied by other children			
Gets on better with adults than with other children			
Considerate of other people's feelings			
Shares readily with other children (treats, toys, pencils etc.)			
Helpful if someone is hurt, upset or feeling ill			
Kind to younger children			
Often volunteers to help others (parents, teachers, other children)			

Signature .....
Date……

Thank you very much for your help

APPENDIX I: Guess the Alien Example

APPENDIX J: Ethics Application and Approval

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:	Méabh Foley
2.2	Your supervisor's name:	Matthew Jones Chesters
2.3	Name(s) of additional UEL supervisors:	Matthew Boardman
		3rd supervisor (if applicable)
2.4	Title of your programme:	Doctorate in Clinical Psychology
2.5	UEL assignment submission date:	Initial submission date
		Re-sit date (if applicable)

Section 3 – Project Details		
Please give as much detail as necessary for a reviewer to be able to fully understand the nature and purpose of your research.		
3.1	Study title: Please note - If your study requires registration, the title inserted here must be the same as that on PhD Manager	Using a game-like task as an assessment of emotion recognition in children
3.2	Summary of study background and aims (using lay language):	Emotion recognition is an important aspect of a child's sociocognitive development however tests in the area often fail to target the specific construct of emotion recognition. The aim of this

		study is to develop a culturally accessible, engaging and enjoyable psychometric tool of emotion recognition for primary school aged children. This tool will be developmentally appropriate and will have a game-like procedure which it is hoped will function to help maintain the attention of the child and reduce the stress of testing.
3.3	Research question(s):	1) Do a non-clinical sample of children implement similar strategies during assessment with the Alien Quiz? 2) Can normative performance characteristics, such as scores and common patterns of responding, which identify normal variation of emotional recognition be established? 3) Do children engage well with the Alien Quiz as a measure of emotional recognition? 4) Is there an association between performance on the Alien Quiz and real-world social functioning?
3.4	Research design:	A mixed methods cross-sectional and correlational design is proposed. This will allow for examination of strategies and performance characteristics on the Alien Quiz and also the exploration of relationships among variables.
3.5	Participants: Include all relevant information including inclusion and exclusion criteria	Participants will be recruited from primary schools in the London region. Children aged between 6-11 years old who understand verbal English will be included in this study.
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 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: Use of Opt-in consent if school does not agree with opt-out consent strategy. Recruitment Plan C: 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.	
3.7	Measures, materials or equipment: Provide detailed information, e.g., for measures, include scoring instructions, psychometric properties, if freely available, permissions required, etc.	Two subscales of the Strengths and Difficulties Questionnaire (SDQ, Appendix A) will be administered to the class teacher of the participating child. The peer problems subscale will be used to measure participant difficulties in their interactions with their peers; the prosocial subscale will be administered to assess the prosocial resources. The Alien Quiz will be developed in order to assess the participants' ability recognise different facial emotions. The test will consist of a game-like task with non-culturally specific named Aliens arranged on a board with hinged frames. The Aliens will have different characteristics which the participant can ask about in order to determine the target Alien. Half of the characteristics will be physical (ex. tail, legs, colour) the other half will be facial emotional expressions derived from Ekman's six basic emotions. A visual analogue scale (Appendix B) will be used to assess whether the participant enjoyed the task.	
3.8	Data collection: Provide information on how data will be collected from the point of consent to debrief	The testing procedure will be carried out in a quiet room in the participant's school. A brief initial assessment will establish the participant's age and that they can understand verbal English. The participant will then be read a standardised, age appropriate set of instructions for the Alien Quiz. The participant's questions will be recorded using pen and paper. Following completion of the Alien Quiz will be asked to mark, on a visual analogue scale their enjoyment of the task. They will also be asked for verbal feedback on the task. It is anticipated that testing will take no longer than 15 minutes per participant.	
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:	Quantitative data will be analysed using SPSS, correlations will be used to assess the relationship between scores on Alien Quiz and the established measures, where possible controlling for age, sex, culture, English as a first language. Planned contrasts will be used to assess the difference between the enjoyability ratings on the established measures and Alien Quiz also controlling for the above. A content analysis will be used to analyse qualitative feedback on the further development of the game.	

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 a participant code to corresponding 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 anonymised 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 is myself, however, the data may also be looked at by my Director of Studies and could be requested by examiners	
4.6	Which data are of long-term value and will be retained? (e.g., anonymised interview transcripts, anonymised databases)	Anonymised assessment transcripts and the anonymised database of quantitative data will be retained for 3 years.	
4.7	What is the long-term retention plan for this data?	The data will be kept for 3 years following the completion of the research. Following submission of the thesis, data will be retained by my Director of Studies and deleted after 3 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 small risk of completing the research during a pandemic.To minimise risk of infection for the participant, current guidelines will be followed i.e. masks will be worn if requested, the room will be large enough for social distancing and hands and surfaces will be regularly washed/sanitized. The researcher 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 researcher will adhere to the school's process for risk assessments. There is no psychological risk to the participant in this research.		
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 a 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 researcher has received both doses of the vaccine and will be completing lateral flow tests twice a week. The researcher will adhere to the school's process for risk assessments.		
5.3	If you answered yes to either 5.1 and/or 5.2, you will need to complete and include a General Risk Assessment (GRA) form (signed by your supervisor). Please confirm that you have attached a GRA form as an appendix:	YES ☒		
5.4	If necessary, have appropriate support services been identified in material provided to participants?	YES ☒	NO ☐	N/A ☐
5.5	Does the research take place outside the UEL campus?	YES ☒	NO ☐	
	If yes, where?	the research 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	YES ☐		

	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.		
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 ☒	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 ☒	NO ☐
6.4	If you have current DBS clearance, please provide your DBS certificate number:	001778281921	
	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.	mainstream primary schools	
	If yes, written permission is needed from such organisations (i.e., if they are helping you with recruitment and/or data collection, if you are collecting data on their premises, or if you are using any material owned by the institution/organisation). Please confirm that you have attached written permission as an appendix.	YES ☒	
7.2	Additional guidance: Before the research commences, once your ethics application has been approved, please ensure that you provide the organisation with a copy of the final, approved ethics application or approval letter. Please then prepare a version of the consent form for the		

	organisation themselves to sign. You can adapt it by replacing words such as 'my' or 'I' with 'our organisation' or with the title of the organisation. This organisational consent form must be signed before the research can commence. If the organisation has their own ethics committee and review process, a SREC application and approval is still required. Ethics approval from SREC can be gained before approval from another research ethics committee is obtained. However, recruitment and data collection are NOT to commence until your research has been approved by the School and other ethics committee/s.

Section 8 – Declarations		
8.1	Declaration by student. I confirm that I have discussed the ethics and feasibility of this research proposal with my supervisor:	YES ☒
8.2	Student's name: (Typed name acts as a signature)	Meabh Foley
8.3	Student's number:	U2075201
8.4	Date:	07/10/2022
Supervisor's declaration of support is given upon their electronic submission of the application		

Student checklist for appendices – *for student use only*

Documents attached to ethics application	YES	N/A
Study advertisement	☒	☐
Participant Information Sheet (PIS)	☒	☐
Consent Form	☒	☐
Participant Debrief Sheet	☒	☐
Risk Assessment Form	☒	☐
Country-Specific Risk Assessment Form	☐	☒
Permission(s) from an external organisation(s)	☒	☐

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:	Méabh Foley
Course:	Prof Doc Clinical Psychology
Title of proposed study:	Using a game-like task as an assessment of emotion recognition in children

Checklist (Optional)	YES	NO	N/A
Concerns regarding study aims (e.g., ethically/morally questionable, unsuitable topic area for level of study, etc.)	☐	☐	☐
Detailed account of participants, including inclusion and exclusion criteria	☐	☐	☐
Concerns regarding participants/target sample	☐	☐	☐
Detailed account of recruitment strategy	☐	☐	☐
Concerns regarding recruitment strategy	☐	☐	☐
All relevant study materials attached (e.g., freely available questionnaires, interview schedules, tests, etc.)	☐	☐	☐
Study materials (e.g., questionnaires, tests, etc.) are appropriate for target sample	☐	☐	☐
Clear and detailed outline of data collection	☐	☐	☐
Data collection appropriate for target sample	☐	☐	☐
If deception being used, rationale provided, and appropriate steps followed to communicate study aims at a later point	☐	☐	☐
If data collection is not anonymous, appropriate steps taken at later stages to ensure participant anonymity (e.g., data analysis, dissemination, etc.) – anonymisation, pseudonymisation	☐	☐	☐
Concerns regarding data storage (e.g., location, type of data, etc.)	☐	☐	☐
Concerns regarding data sharing (e.g., who will have access and how)	☐	☐	☐
Concerns regarding data retention (e.g., unspecified length of time, unclear why data will be retained/who will have access/where stored)	☐	☐	☐
If required, General Risk Assessment form attached	☐	☐	☐
Any physical/psychological risks/burdens to participants have been sufficiently considered and appropriate attempts will be made to minimise	☐	☐	☐
Any physical/psychological risks to the researcher have been sufficiently considered and appropriate attempts will be made to minimise	☐	☐	☐
If required, Country-Specific Risk Assessment form attached	☐	☐	☐

If required, a DBS or equivalent certificate number/information provided	☐	☐	☐
If required, permissions from recruiting organisations attached (e.g., school, charity organisation, etc.)	☐	☐	☐
All relevant information included in the participant information sheet (PIS)	☐	☐	☐
Information in the PIS is study specific	☐	☐	☐
Language used in the PIS is appropriate for the target audience	☐	☐	☐
All issues specific to the study are covered in the consent form	☐	☐	☐
Language used in the consent form is appropriate for the target audience	☐	☐	☐
All necessary information included in the participant debrief sheet	☐	☐	☐
Language used in the debrief sheet is appropriate for the target audience	☐	☐	☐
Study advertisement included	☐	☐	☐
Content of study advertisement is appropriate (e.g., researcher's personal contact details are not shared, appropriate language/visual material used, etc.)	☐	☐	☐

Decision options

APPROVED	Ethics approval for the above-named research study has been granted from the date of approval (see end of this notice), to the date it is submitted for assessment.
APPROVED - BUT MINOR AMENDMENTS ARE REQUIRED BEFORE THE RESEARCH COMMENCES	In this circumstance, the student must confirm with their supervisor that all minor amendments have been made before the research commences. Students are to do this by filling in the confirmation box at the end of this form once all amendments have been attended to and emailing a copy of this decision notice to the supervisor. The supervisor will then forward the student's confirmation to the School for its records. Minor amendments guidance: typically involve clarifying/amending information presented to participants (e.g., in the PIS, instructions), further detailing of how data will be securely handled/stored, and/or ensuring consistency in information presented across materials.
NOT APPROVED - MAJOR AMENDMENTS AND RE-SUBMISSION REQUIRED	In this circumstance, a revised ethics application must be submitted and approved before any research takes place. The revised application will be reviewed by the same reviewer. If in doubt, students should ask their supervisor for support in revising their ethics application. Major amendments guidance: typically insufficient information has been provided, insufficient consideration given to several key aspects, there are serious concerns regarding any aspect of the project, and/or serious concerns in the candidate's ability to ethically, safely and sensitively execute the study.

Decision on the above-named proposed research study

Please indicate the decision:	**APPROVED - MINOR AMENDMENTS ARE REQUIRED BEFORE THE RESEARCH COMMENCES**

Minor amendments

Please clearly detail the amendments the student is required to make
Section 3.6 – must specific exact recruitment strategy, how will you gain access to the parents to obtain consent. How will you identify the school/s to approach and what if they do not permit your research. **Section 5.1. This section is more around physical or psychological risks posed by the research. Be clear if the research poses any risks.** **The details provided a good re covid measures however is it realistic to get 6 -11 year olds to ear masks?**

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:	26/10/2022

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)	Meabh Foley
Student number:	U2075201
Date:	27/10/2022

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

APPENDIX K: Research Flier

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:

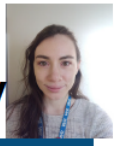
alien.game@uel.ac.uk

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 L: Organisation Invitation letter

ORANISATION INVITATION LETTER

Using a game-like task as an assessment of emotion recognition in children.

Contact Person: Meabh Foley
Email:m.foley2001@uel.ac.uk

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 am I?

My name is Meabh Foley, I am a Trainee Clinical Psychologist. This study is being conducted as part of my Professional Doctorate in Clinical Psychology at the University of East London.

What is the research?

I am conducting research into improving tests of social cognition for children. Social cognition includes children's ability to process, store, and apply information about other people and social situations. The aim of this study is to assess whether a newly developed game can successfully measure social cognition in children in a more engaging and accessible manner than tests that are currently available. I am also looking to obtain feedback from children on how the game could be further developed.

Current tests of social cognition are often limited by time, cultural norms and language. This newly developed game hopes to address some of these limitations and if children do find it more engaging it could help us measure these skills more accurately.

My 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 I am looking for to help me explore my research topic. I am looking to involve children aged 6 to 11.

The children who agree to participate will not be judged or personally analysed in any way and will be treated with respect.

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 attend a session with myself, 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. They will then be asked to complete a new game developed to measure the child's ability to store, and apply information about other people and social situations, after which they will be asked to rate their level of enjoyment for the task. Children's verbal responses during the game-like assessment will be recorded using pen and paper. The session should take about 20 minutes and will take place at your location. We would also ask the child's teacher to fill in a brief questionnaire about the child's ability to apply information about other people and social situations. The aim of this is to find out whether the measures are related to real-life strengths and/or difficulties.

I 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 my research topic.

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 my own safety, social distancing will be maintained at all times, I will wear a mask and sanitizing of hands and equipment will be completed regularly.

What will happen to the information provided?

What I will do with the material children provide will involve anonymously storing all data on a personal drive, only I 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 myself and my 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 November 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 November 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 my research or have any questions or concerns, please do not hesitate to contact me. My email address is m.foley2001@uel.ac.uk.

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

or

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

or

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

Thank you for taking the time to read this information sheet

APPENDIX M: Head Teacher's *In Loco Parentis* Form.

UNIVERSITY OF EAST LONDON

Using a game-like task as an assessment of emotion recognition in children

Head Teacher's *In 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 (1 week) 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 N: Guardian Information Sheet and Opt-out Consent Form

PARENT/GUARDIAN PARTICIPANT INFORMATION SHEET

Using a game-like task as an assessment of emotion recognition in children

Contact person: Méabh Foley
Email: m.foley2001@uel.ac.uk

Your child is being invited to participate in a research study. Before you decide whether to take part or not, please carefully read through the following information which outlines what your 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 me on the above email.

Who am I?

My name is Méabh Foley, I am a Trainee Clinical Psychologist. This study is being conducted as part of my Professional Doctorate in Clinical Psychology at the University of East London

What is the purpose of the research?

I am conducting research into improving tests of social cognition for children. Social cognition includes children's ability to process, store, and apply information about other people and social situations. The aim of this study is to assess whether a newly developed game can successfully measure social cognition in children in a more engaging and accessible manner than tests that are currently available. I am also looking to obtain feedback from children on how the game could be further developed.

Current tests of social cognition are often limited by time, cultural norms and language. This newly developed game hopes to address some of these limitations

and if children do find it more engaging it could help us measure these skills more accurately.

My 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 have I been invited to take part?

To address the study aims, I am inviting children aged 6-11 to take part in my research. Your child will not be judged or personally analysed in any way and will be treated with respect.

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

What will I be asked to do if I agree to take part?

If you agree for your child to participate your child will be asked to attend a session with myself, where they will be asked some background questions, such as their date of birth, gender, ethnicity, country of birth, first language, main language spoken at home. They will then be asked to complete a new game developed to measure the child's ability to store, and apply information about other people and social situations, after which they will be asked to rate their level of enjoyment for the task. Your child's verbal responses during the game-like assessment will be recorded using pen and paper. Your child will be asked for feedback on how the game could be further developed. The session should take about 20 minutes and will take place in a quiet room at their school. We would also ask your child's teacher to fill in a brief questionnaire about the child's ability to store, and apply information about other people and social situations. The aim of this is to find out whether the measures are related to real-life strengths and/or difficulties.

I will not be able to pay your child for participating in my research, but their participation would be very valuable in helping to develop knowledge and understanding of my research topic.

Can I change my mind?

Your child is free to withdraw from the research study at any time without explanation, disadvantage or consequence. Separately, your child may also request to withdraw their data even after they have participated data, provided that this

request is made before February 2023 (after which point the data analysis will begin, and withdrawal will not be possible). You can request withdrawal of data by letting Méabh know via the email address at the top of this letter

Are there any disadvantages to taking part?

No disadvantages to taking part have been identified.

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

Your child's privacy and safety will be respected at all times. I 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. Anonymised data will be accessed only by this researcher and research supervisor Dr Matthew Jones Chesters. You and your child's consent form will be stored securely, using UEL MS Teams and OneDrive, and separately from the rest of the data. Your child's answers will be securely stored on MS teams, they will be deleted following data analysis. Consent forms will be destroyed following completion of the research. Anonymised research data will be securely stored by my supervisor, Dr Matthew Jones Chesters, for a maximum of 3 years, following which all data will be deleted.

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

What will happen to the results of the research?

Summaries of the data collected will be available in the write-up as a thesis and submitted for assessment. The thesis will be publicly available on UEL's online. Findings will also be disseminated to a range of audiences (e.g., academics, clinicians, public, etc.) through journal articles and conference presentations. 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 or the school they attend.

You will be given the option to receive a summary of the research findings once the study has been completed for which relevant contact details will need to be provided.

Who has reviewed the research?
My 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 my research or have any questions or concerns, please do not hesitate to contact me. My email address is m.foley2001@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

UNIVERSITY OF EAST LONDON

PARENTAL/GUARDIAN CONSENT OPT-OUT FORM

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

Using a game-like task as an assessment of emotion recognition 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 me on **m.foley2001@uel.ac.uk.**

Your child's participation in the study is voluntary and you can withdraw them at any time before **February 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 **m.foley2001@uel.ac.uk.**

If you do not want your child to take part in the study, (1) check the box below, (2) sign the form and date it, and (3) return it to the school within 3 working days. You can contact me

via **m.foley2001@uel.ac.uk** or speak with the school team if you have any questions. Thank you.

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

Child's name (please print) _____

Child's age group_____

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

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

Parent / guardian signature_____

Date_____