The Utility of Virtual Reality in Interventions for Autism Spectrum Disorder: A Systematic Review

Victoria Caroline Brattan

Submitted in accordance with

the requirements for the degree of

Doctor of Clinical Psychology (D. Clin. Psychol.)

The University of Leeds

School of Medicine

Academic Unit of Psychiatry and Behavioural Sciences

October 2019

AUTHOR'S DECLARATION

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

ACKNOWLEDGEMENTS

A very big thank you to my thesis supervisor, Dr. Faisal Mushtaq, for his infinite patience, support and encouragement over this time of training.

Also a huge thank you to my clinical and academic tutors, Dr. Jan Hughes and Dr. Gary Latchford, for their continued support, understanding, time and compassion over the course of the DClin training.

A sincere thank you to my friends on the course, Rhian, Kasia, Saafi, Dani, Daisy and Rion for your listening ear and compassion, as well as the fun times. And a massive thank you to my long-time friends from other parts of life, Laura, Marta and Parisa. Though I don't see you often, I know you are always there. Without you I wouldn't have got through this.

Most importantly of all, thank you to Mum and Dad, without you I would not have finished this training and would not have achieved or done all that I have so far. Thank you for being the biggest and best teachers in life, for always grounding me, putting things into perspective, and reminding me of what's most important in life – family, friends, health, and most importantly, love.

ABSTRACT

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterised by impairments in social communication and interaction. Prevalence rates of anxiety are higher in this population compared to the general population. Anxiety and autistic traits can seriously impede an individual's capacity to function in the social world. Current psychosocial interventions for ASD individuals are aimed to develop skills in interaction and communication or to address anxiety. The development and use of virtual reality for clinical interventions is on the rise, and its potential benefits for ASD individuals are numerous. As yet, little is known about the utility of VR-based interventions for ASD.

We conducted a systematic review of randomised and non-randomised studies that employ VR for intervention in the core deficits of ASD and/or anxiety, and which report pre- post-intervention data or change over time. Twenty-four studies met the inclusion criteria for the review. A quality assessment of included studies was conducted to evaluate their risk of bias. The review yielded five randomised controlled studies and 19 non-randomised or case studies. Studies employed VR in its variety of forms, including head-mounted displays, desktop VR, and cave environments. A variety of interventions were employed, including CBT, scaffolded hierarchical learning, and social cognition training.

Findings from the review suggest that VR-based interventions for ASD individuals are feasible and demonstrate effectiveness in the development of affect recognition and emotion regulation skills, as well as for job interviewing skills. Additionally, studies demonstrate its promise for development of communication and conversational skills. Further research is required of higher quality to determine the efficacy and effectiveness of studies in this and other areas. In particular, it is important that studies progress from exploratory use of VR toward more theory and evidence-informed intervention protocols for ASD individuals. Additionally, follow-up research studies of the impact of intervention on individuals' daily lives is also necessary to determine the generalisability of skills developed in VR, and real-life impact.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The Utility of Virtual Reality in Interventions for Autism Spectrum Disorder: A Systematic Review

1. INTRODUCTION AND LITERATURE REVIEW

1.1. What Is Autism Spectrum Disorder? Diagnostic criteria

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, the diagnosis of which is reliant on the observation, and reported observation, of impairments within the domains of (i) social communication and social interaction across multiple contexts, and (ii) restricted, repetitive and stereotyped behaviours (the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition, DSM-V, American Psychiatric Association, 2012). The diagnostic criteria for autism are used as a framework for the discussion and decision-making within a multidisciplinary team (MDT).

To be given a diagnosis of ASD, an individual must meet criteria A, B, C, and D:

A. Persistent deficits in social communication and social interaction across multiple contexts, and manifest by all three of the following:

1. Deficits in social-emotional reciprocity
2. Deficits in nonverbal communicative behaviours used for social interaction
3. Deficits in developing, maintaining, and understanding relationships

B. Restricted, repetitive patterns of behaviour, interests, or activities, as manifested by at least two of the following:

1. Stereotyped or repetitive motor movements
2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behaviour
3. Highly restricted, fixated interests that are abnormal in intensity, or
4. Hyper- or hypo-reactivity to sensory input, or unusual interest in sensory aspects of the environment

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay.

For a complete outline of DSM-V diagnostic criteria and required evidence within each of the domains, please see Appendix A.

The DSM does not distinguish between subtypes of autism, and refers to them all under the umbrella term of autism spectrum disorder (ASD). The International Statistical Classification of Diseases and Related Health Problems, Eleventh Edition (ICD-11, World Health Organisation (WHO), 2018) is now in line with the diagnostic criteria of the DSM. As such, the terms autism and ASD are used interchangeably throughout the sections that follow.

1.2. Anxiety in ASD

There is a body of evidence that shows a high prevalence of anxiety and anxiety disorders in individuals with ASD and that people with ASD are at higher risk of developing anxiety than the general population (Leyfer et al., 2006; MacNeil, Lopes, & Minnes, 2009; Simonoff et al., 2008; van Steensel, Bögels, & Perrin, 2011; White, Oswald, Ollendick, & Scahill, 2009).

Van Steensel et al. (2011) conducted a meta-analysis of literature examining anxiety in children and adolescents with ASD to determine which of the anxiety specific disorders are most prevalent in ASD. Thirty-one studies were identified that used standardised assessments of anxiety in ASD, that together involved over two thousand young people below the age of eighteen with ASD without learning disability. Across these studies, nearly 40% of the young people with ASD had at least one anxiety disorder (according to the DSM-IV) or scored above clinical cut-off for anxiety in general. The most prevalent anxiety disorder was specific phobia (29.8%), followed by OCD (17.4%), and social anxiety disorder (16.6%), generalized anxiety disorder (nearly 15%), separation anxiety disorder in nearly 9%, and panic disorder in nearly 2%.

Bradley, Summers, Wood, and Bryson (2004) conducted a study comparing anxiety in twelve children with ASD and learning disability to children with learning disability but without ASD. On the anxiety subscale of the revised Diagnostic Assessment for the Severely Handicapped (DASH-

II, Matson, 1995), 42% of the sample with ASD reached clinical significance compared to none of the individuals without ASD. Similarly, Gillott and Standen (2007) examined anxiety in adults with ASD and intellectual disabilities compared to adults with intellectual disabilities but without ASD. They found that the autistic adults had significantly higher overall anxiety than the control group as measured by the total score on a modified version of the Spence Children's Anxiety Scale–parent (Spence, 1999). With respect to specific anxiety disorders, the group with autism had significantly higher scores on agoraphobia, separation anxiety, obsessive-compulsive disorder, and generalized anxiety disorder.

With respect to levels of anxiety experienced, studies suggest that young people with ASD and anxiety, and neurotypical young people who are clinically anxious experience similar levels of anxiety (Farrugia & Hudson, 2006; Russell & Sofronoff, 2005). Farrugia and Hudson (2006) compared anxiety symptoms in adolescents with ASD to those without ASD but diagnosed with an anxiety disorder, as well as to non-clinical adolescents. Based on both self-report and parent ratings of anxiety, adolescents with ASD scored significantly higher than the non-clinical young people, and anxiety ratings were equivalent in the ASD group to those in the anxiety disorders group.

In summary, children and adolescents with ASD have been shown to have same levels of anxiety as clinically anxious typically developing children, but higher anxiety ratings compared to non-clinical groups. Of importance, the results of the above studies demonstrate substantial comorbidity for anxiety in children, adolescents, and adults with autistic spectrum disorder. Given that 40% of young people with ASD are estimated to have an anxiety disorder of any kind, yet only 2.2. – 27% of typically developing children are estimated to have an anxiety disorder (see Costello, Egger, & Angold (2005) for a review), rates of anxiety in ASD are far above the norm in the neurotypical population and as such is a significant concern.

Research suggests that there are processes and symptoms that are common to anxiety in individuals both with and without ASD and which conform to DSM criteria, as well as processes that are atypical and inherent (even potentially unique) to ASD that may not conform to traditional diagnoses of anxiety (Kerns et al., 2018; Ollendick & White, 2012). Physiologically, the experiences of anxiety are the same in neurotypicals and individuals with ASD (Ollendick & White, 2012). Other common anxiety processes to ASD and neurotypicals include negative-biased information processing and automatic or unhelpful thoughts (Ollendick & White, 2012). The distinctions lie, in part, in the types of fears and anxieties that individuals with ASD present

with. For instance, fears of particular sounds and noises (e.g. hand dryers, vacuum cleaners), as well as unusual and highly specific phobias and worries (for instance, fear of toilets or bridges), as opposed to more generalised anxieties, are a common feature of the presentation of anxiety in ASD, and may be related to atypical sensory experiences (Kerns et al., 2014). Studies also suggest that anxiety in ASD may present as nervousness, social distress and social avoidance without (or with limited) awareness or concern for social rejection (Leyfer et al., 2006; Muris, Steerneman, Merckelbach, Holdrinet, & Meesters, 1998), as well as fewer fears of social evaluation and self-consciousness (see Kerns & Kendall 2012; White et al. 2009). Such a presentation thus does not conform to the traditional diagnosis of social phobia (Leyfer et al. 2006; Muris et al. 1998).

Further distinctions in anxiety presentation lie in how individuals are able to read and interpret their internal physical and emotional states. Individuals with ASD are less likely to be able to identify the emotion being experienced in a particular situation and understand their physical response compared to neurotypical individuals. In turn, they may be less able to verbally express their discomfort. This can lead to manifestations of anxiety in individuals with ASD that may differ from typical displays of anxious states. For example, anxiety may manifest in children in ways that include tantrums, aggression, self-injury, non-compliance, and some repetitive behaviours as a means to reduce feelings of anxiety. In adults, anxiety may manifest in more subtle ways and can include social avoidance, negative or sometimes hostile interpersonal exchanges, sensory defensiveness, and rigidity or adherence to routines (Lecavalier et al., 2014; Ollendick & White, 2012).

Consistent with the Ollendick and White (2012) model, Kerns et al., (2014) found that traditional anxiety in ASD was associated with processes shared with neurotypicals (anxious cognitive style; hypersensitivity), whereas only atypical anxiety symptoms were also associated with ASD symptoms. Kerns et al (2014) demonstrated that in a sample of 59 ASD youth, forty-six percent displayed symptoms of anxiety that were not consistent with any DSM psychological disorders. Rather, symptoms reflected a more diffuse form of anxiety not currently captured in the DSM. In general, it would seem that symptoms of anxiety in ASD manifest as exacerbated and functionally-impairing anxiety which converges with features of ASD.

If anxiety presents as quite different in ASD compared to the general population, this has implications for the assessment and measurement of anxiety, which would require modification for the ASD population. For instance, when considering self-report measures of anxiety, the ability to identify internal states is required as well as adequate expressive language skills and

the ability to communicate with emotional language. As such, individuals with ASD may struggle with self-report measures of anxiety. Similarly, parent-report measures of anxiety rely on parental awareness of the influence of anxiety on their child's behaviour, as well as verbal expressions of anxieties by the child.

As current research indicates individuals with ASD may experience anxiety in an atypical and more diffuse way that traditional anxiety symptoms, measures of anxiety would need to reflect this. Idiosyncratic measures, specific to an individual's fears/phobias for instance, should be considered for the evaluation of anxiety in ASD. Additionally, measures of anxiety in ASD may benefit from using open-ended questions, rather than standardised lists of fears as used for the general population (Mayes et al., 2013). Consideration should also be given to the convergence of anxiety and ASD traits and atypical manifestations of anxiety. Some anxiety symptoms and ASD traits can overlap and be difficult to differentiate (MacNeil et al., 2009; Tsai, 2006). For example, it can be difficult to differentiate between compulsions (repetitive actions used to reduce anxiety) which are central to OCD (APA, 2000), and the repetitive and ritualistic behaviour traits inherent to ASD (Cath, Ran, Smit, Van Balkom, & Comijs, 2007; Gjevik, Eldevik, Fjæran-Granum, & Sponheim, 2011; Rodgers, Glod, Connolly, & McConachie, 2012), but which may also in fact be an atypical manifestation of anxiety (such as repetitive movements or sensory seeking behaviours as a means to self-sooth).

This overlap could therefore raise questions about the reported prevalence rates of anxiety in the ASD population; that is, rates of anxiety may be underreported if they are masked as ASD traits (Simonoff et al., 2008). Current prevalence rates may be a consequence of the diagnostic overlap, particularly in the case of OCD and social phobia, as well as the use of standardised measures of anxiety that have not generally been developed for the ASD population. As such, the assessment of anxiety in some cases may be argued to be akin to assessing severity of ASD traits. However, studies that used ASD adapted measurements (e.g., excluding several items, Kuusikko et al., 2008; Sukhodolsky et al., 2008; or training interviewers to distinguish anxiety from ASD, Simonoff et al., 2008) reported prevalence rates of anxiety disorders ranging from 31.5% to 50.0% in the ASD population, in accordance with van Steensal et al.'s (2011) estimated rates. Furthermore, the van Steensal et al. (2011) meta-analysis found that children with PDD-NOS had higher prevalence rates of anxiety disorders, suggesting that the less severe ASD subtype (as compared to Asperger's syndrome and autistic disorder) have the highest anxiety scores, which is contradictory to the notion that assessing anxiety is the same as assessing severity of ASD.

The difficulties with differentiating between anxiety symptoms and features of autism is to be expected. There are features of the ASD phenotype that can generate increased anxiety (e.g. difficulties with understanding what to do or say in social interactions can create increased levels of anxiety in social situations). At the same time, anxiety can exacerbate autistic-like traits (e.g. repetitive behaviours used to reduce anxiety). Gillott & Standen (2007) conducted a study to examine levels and sources of stress and anxiety in a group of adults with ASD and learning disabilities compared to a group with learning disabilities without ASD. In addition to finding that the group with autism had significantly higher overall anxiety as well as specific anxiety disorders, they found that the sources of stress that correlated with high anxiety levels for the autism group related specifically to autistic traits. That is, they report the sources of stress for the ASD group included the ability to cope with change, anticipation, sensory stimuli and unpleasant events, which relate to the autistic traits of resistance to change and adherence to routines, hyper/hypo-reactivity to sensory input, and emotion regulation difficulties respectively. It could be argued, therefore, that anxiety symptoms and autistic traits are inextricably linked. Many traits of ASD can engender anxiety, particularly in high functioning individuals who have greater awareness of their social impairments and differences (White et al., 2014). Anxiety is such a problem in the ASD population that some researchers posit that it is a core feature of ASD itself, and may form a feature of the broader phenotype (MacNeil et al., 2009). Yet anxiety itself, whether considered as a comorbid disorder or as an epiphenomenon of autistic traits, can contribute to functional impairment over and above the deficits of ASD (Matson & Nebel-Schwalm, 2007).

ASD is a life-long condition. As individuals mature from childhood into adolescence and into emerging adulthood they increasingly hope to become more independent, leave the parental home, aim to study further or gain employment, and enter into relationships. These everyday aspects of life can be particularly difficult for individuals with ASD as they enter adulthood as their home base changes from the familiarity once given by the family home; independent living is challenged by greater unpredictability in crowded, noisy environments in supermarkets, coffee houses and public transport, where they no longer have the support or scaffolding of their parents; and work or education can be hindered and stressful as the structure and support of early school life is removed. At university, students with ASD are known to have higher than average drop-out rates, experience greater academic difficulty, experience greater levels of social isolation as well as emotional distress and mental health problems including anxiety and depression (Smith, Maenner, & Seltzer, 2012; White, Ollendick, & Bray, 2014; White et al., 2016). The impact on life of ASD traits and as individuals enter into adulthood is therefore far-reaching

as it limits individuals' likelihood to develop independence, likelihood of engaging in social encounters, thus engendering social isolation, and limiting the possibility of individuals' capacity to reach their full potential in work or education.

1.3 Psychological Interventions

Despite the inter-relatedness of ASD traits and anxiety symptoms, the interventions currently offered to ASD individuals are generally aimed to tackle either (i) core deficits of social communication and interaction, or (ii) anxiety. Though individuals with ASD also experience other comorbidities, such as depression, the vast majority of intervention research focusses on anxiety (Walters, Loades, & Russell, 2016).

1.3.1 Interventions for core deficits of ASD

Treatments to address core deficits have focussed primarily on addressing impairments in social communication and interaction. Research over the last ten years has provided evidence in support of certain models and methods, in particular psychosocial intervention and social skills training packages, to address these issues in individuals with ASD.

With respect to psychosocial interventions for the core social communication and interaction deficits, social skills training (SST), or Social Cognition Training (SCT) is the most widely used method to address social communication and interaction problems in ASD (Lerner, White, & McPartland, 2012). Such programmes are based on the premise that prosocial behaviours can be fostered through structured learning opportunities and practice in a 'safe' structured environment (in-office) and then generalised to other contexts through out-of-office learning opportunities. These interventions generally involve facilitated practice with role-play to develop understanding of social conventions and skills. They include a very broad array of methods, including social stories (Sansosti & Powell-Smith, 2008), video-modelling (Sansosti & Powell-Smith, 2008), pivotal response training (Koegel, Koegel, Vernon, & Brookman-Frazee, 2010), and peer-mediated training and intervention (Laugeson, Frankel, Mogil, & Dillon, 2009) amongst others, but these topics are beyond the remit of this thesis.

Cognitive Behavioural Therapy (CBT) is the predominant psychotherapy model that has been used with individuals with ASD to address social cognition and interaction difficulties (Kreslins, Robertson, & Melville, 2015). CBT is based on the theory that thoughts, feelings, physiological reactions and actions interrelate to maintain maladaptive and unwanted 'symptoms' or

behaviours. Therapy focuses on cognitions, feelings and actions that are accessible to the individual in order to address problematic responses and facilitate change. CBT tends to be very structured and implements the use of an agenda to keep therapeutic sessions on target and strategy to achieve realistic goals. Recent studies have examined the use of CBT to improve social communication and social interaction difficulties through the identification of unhelpful thought processes and behaviour patterns (Wood, Fujii, Renno, 2011; Klinger & Williams, 2009; Bauminger, 2002, 2007) under the CBT premise that social perception processes and social competence can be taught cognitively and can influence behaviour (Hart & Morgan, 1993).

Some researchers have evaluated the effectiveness of combining the CBT approach with social skills training in school settings to address core deficits of ASD (Bauminger, 2002, 2007). Bauminger (2002) examined this approach for fifteen school children, and reported positive improvements in socio-emotional functioning and interactions following CBT-framed SST delivered within individual children's schools. However, the intervention was delivered by a teacher who reported on progress and an observer who was not blind to the study's aims. Bauminger (2007) thus addressed this limitation using more objective evaluators in both (i) an individual, and (ii) a group intervention programme. The study evaluated the SST conducted within the CBT framework which aimed to enhance socio-emotional understanding, and dyadic as well as group social interaction. It expanded on the previous study in that it examined (a) reports of children's overall progress in social skills by objective observers; (b) intervention effects 4 months post-treatment on children's social interaction capabilities; and (c) change in children's self-reported loneliness and self-perception.

The intervention involved three main domains: (i) instruction in social concepts, such as the understanding of friendship; (ii) psychoeducation on emotions (related to the four basic emotions of sadness, happiness, fear, and anger); and (iii) interpersonal problem solving, with a focus on 13 core social objectives, such as initiating a conversation. The results of the two-year programme demonstrated significant positive improvements in the children's dyadic interactions and social cognition abilities (emotion recognition and understanding, and social problem solving). In general, reviews suggest that SSTs are broadly supported for improving social interaction abilities in ASD populations (Reichow & Volkmar, 2010; Wang & Spillan, 2009), but not entirely consistently as it may be limited by age and developmental levels (Bellini, Peters, Benner, & Hopf, 2007).

Lerner, White, and McPartland (2012) examine why and how CBT and SST focussed intervention might work. They theorise the potential mechanisms underpinning the interventions are common to many treatment modalities. Lerner et al. (2012) suggest particular mechanisms that are potentially influential (in terms of moderators and mediators) in engendering change in people with ASD. These include behavioural modification, the therapeutic relationship, social knowledge, social motivation, social information processing, executive functioning, and internalised comorbidities (e.g. depression and anxiety). They point out the impact that comorbidities, including anxiety, can have for an individual with ASD, in this case, on the extent of effectiveness of an intervention. As such it is important to address this trait which not only potentially engenders greater ASD behaviours and characteristics, but can also impede effectiveness of intervention.

1.3.2 Interventions for anxiety in ASD

Characteristics and traits of individuals with ASD can complicate intervention for ASD traits and its comorbidities, and in turn treatment efficacy. For example, individuals with ASD often have a reduced awareness and acknowledgement of thoughts and feelings, both in themselves and others (DSM-V, 2013), which can hinder introspection. Additionally, language and social skill deficits inherent to ASD can impede the development of the therapeutic relationship (Lang, Regester, Lauderdale, Ashbaugh, & Haring, 2010), and other features of ASD, such as concrete thinking, potentially impact the efficacy of standard treatment (Lickel, Maclean, Blakeley-Smith, & Hepburn, 2012). This has led to a debate about whether CBT is appropriate for this population (Chalfant, Rapee, & Carroll, 2007; Lickel et al., 2012).

Attwood (2004) suggests additions to intervention to benefit people with ASD. These include: (a) increasing the use of visual aids, (b) using social stories to explain complex scenarios and expectations, (c) associating emotions with tangible objects (e.g. making a scrapbook of relevant pictures, creating drawings of feelings and thoughts), (d) increased emphasis on coping strategies that do not require the use of abstract language (e.g. relaxation strategies), (e) use of alternative communication modes (e.g. internet chat), (f) embedding perseverative interest topics into sessions and (g) increasing the focus on teaching social skills. Other researchers have made similar suggestions for modifying CBT for individuals with ASD (e.g. Anderson & Morris, 2006). Modifications are largely focused on the structure and mode of delivery of CBT rather than the content of the intervention and point to the need to reduce, or simplify, cognitive

components. This reduction in cognitive elements of intervention further emphasises the questionability of CBT as the most appropriate method for ASD clients.

Nevertheless, CBT remains the most employed intervention approach that has been empirically evaluated by randomised controlled trial (RCT) for anxiety in individuals with ASD (Kreslins et al., 2015), and is the recommended treatment to manage anxiety in young people with ASD if they have the necessary verbal and cognitive abilities (NICE, 2014). Very few published studies that examine the efficacy of CBT for ASD have not been modified to some degree for the population (Walters et al., 2016). Adaptation and extensions to the general CBT approach are made in order to increase the effectiveness of treatment with this population. Several studies have made some of the adaptations as suggested by Attwood (2004), including the use of an increased number and quality of visual aides to teach concepts (Chalfant et al., 2007; Reaven et al., 2009; Reaven & Hepburn, 2003; Sze & Wood, 2007, 2008). Visual aides in these cases have involved cardboard cut-outs of representations of coping strategies and photographs of anxiety-evoking stimuli. Some studies have also integrated the use of the individual's stereotyped or perseverative interests as a means to increase engagement and motivation (Reaven et al., 2009; Sze & Wood, 2007, 2008; Wood et al., 2009). For example, Sze and Wood (2007) reported the case of a young girl with an intense special interest in films from the 1980s, including Indiana Jones. Aspects of the intervention, including cognitive restructuring, were re-framed to incorporate Indiana Jones in a variety of ways. This adaptation was successful in maintaining the young girl's motivation to engage in the treatment and was considered instrumental in eliminating the young girl's anxieties about the Iraq war.

Within the nine studies included in the systematic review conducted by Lang, Regester, Lauderdale, Ashbaugh, and Haring (2010) that examined modified CBT for anxiety in ASD, each of them reported a reduction in anxiety according to at least one outcome measure. However, not all of the modified CBT intervention studies reported significant decreases in anxiety across both the client and other-reported measures. Both Reaven et al. (2009) and Wood et al. (2009) reported significant decreases in children's anxiety according to parent ratings, but no decrease according to the child's own self-report.

In a later systematic review and meta-analysis by Weston, Hodgekins, & Langdon (2016), 48 studies were identified that used CBT as the treatment for affective disorders in ASD or to treat core deficits. Of those studies that assessed the effectiveness of CBT for affective disorders, 19 out of 24 targeted anxiety. Eleven of these studies reported on self-report measures. The

random effects meta-analysis of these revealed a non-significant small to medium effect size, which was further reduced when an outlier study was removed (g=.08). Informant-based outcome measures revealed a significant medium effect size (g=.49), as did clinician-rated outcome measures (g=.60). In the studies that employed CBT as an intervention for core deficits of ASD, a small-to-medium non-significant effect size was revealed with self-report measures (g=.25), a significant small-to-medium effect size with informant-report measures (g=0.48), a significant medium effect size for clinician-report measures (g=0.65), and a significant small-to-medium effect size, for task-based measures g=0.35. Thus, supporting the findings of Lang's systematic review, Weston et al.'s (2016) meta-analysis highlights the discrepancies between effectiveness of CBT treatment according to the client's own self-report and parent or clinician report. If the client's own perception of their anxieties are not reduced, the effectiveness of the treatment must be queried.

Lang et al.'s (2010) review shines a spotlight on CBT's mechanism of action and calls it into question for the ASD population. CBT is assumed to influence levels of anxiety predominantly by modifying dysfunctional cognitive patterns, which relies substantially on introspection. However, many of the adaptations made to CBT for individuals with ASD de-emphasise this introspection, and emphasise the use of practical skill building (Lang et al., 2010). Thus, the adaptations made to CBT for ASD suggest a more behavioural approach, and as such, the utility of CBT and the tendency of practitioners and researchers alike to focus on CBT as an approach for anxiety intervention in this population is questioned. There is a major scarcity of research examining the effectiveness of more behavioural psychological intervention approaches, such as Acceptance and Commitment Therapy (ACT; Harris, 2009) for ASD.

1.4 Use of technology in interventions

Computer-assisted and technology-based interventions have been suggested to hold great promise for therapeutic intervention in the general population as well as in the ASD population. Use of computer-based interventions is argued to increase ease of access to therapeutic skills and techniques (Insel & Sahakian, 2012). It is also argued to increase clinical efficacy and improve cost-efficiency (Insel & Sahakian, 2012), as a program can be used repetitively. In the ASD population, it has also been suggested as a potentially powerful tool as it can circumvent the issues around the limited number of clinicians who are trained in psychotherapeutic modifications specific to ASD (White, 2012) and thus increase access to intervention strategies. Additionally, computer-based interventions have high predictability, and allow an individual to

work at their own pace and repeat steps as many times as they wish, without fatiguing the clinician. For many individuals with ASD, the use of computer technology leverages this often common interest to increase their motivation, engagement and investment in the intervention (Parsons & Mitchell, 2002).

In more recent decades, the advancement of technology has led two-dimensional computer-generated software into a whole new domain, that of the more sophisticated, immersive, three-dimensional digital world of virtual reality (VR). Virtual Reality can come in many forms. In its most sophisticated of forms, with the use of head-mounted displays (HMDs), an environment is rendered around the player and the player's own movements are tracked and digitally displayed within the digital environment. The virtual environment may be a digital recreation of a real-world scenario or an entirely novel one. The idea is that the visual cues of the real world are replaced with digital ones. The most compelling of VR experiences are those that block real world sensory experiences and immerse the senses in the virtual environment (Biocca and Levy, 1995; Witmer& Singer, 1998). Virtual environments provide a substitute reality in which one can interact in the here and now with objects, people, and environments, in a novel and controlled manner, thus allowing for the study of psychological phenomena such as action prediction mechanisms in sporting manoeuvres (Diaz, Cooper, Rothkopf, & Hayhoe, 2013), and social interaction, amongst other areas (Loomis, Blascovich, & Beall, 1999).

The increased use of technology including VR in mental health intervention has been predicted by clinicians and organisational bodies alike. In a study using Delphi methodology, a panel of 62 psychotherapy experts answered questions relating to how future changes were likely to impact on psychotherapy and psychologists. The study identified the prediction that the use of VR and computerised psychotherapies were likely to increase and impact the way psychotherapy in delivered (Norcross, Hedges, & Prochaska, 2002). Moreover, the NHS published a report in 2018, 'The Digital Future of Mental Heathcare and its Workforce' (Foley & Woollard, 2018) in support of the TOPOL Review (Topol, 2019), in which it posits how different technologies will transform mental healthcare over the next 20 years. Among its predictions of key technologies due to impact mental healthcare and its workforce, it identifies VR as due to make an impact within the next three to five years, citing evidence for the use of VR in the management of phobic and psychotic disorders. It additionally identifies VR as a new modality for communication at a distance that can be used to aid isolated or hard-to-reach people, as well as patients with severe anxiety or paranoid disorders that may find face-to-face contact difficult at the start.

It should be noted, however, that the need for involvement of a real person to facilitate such intervention should not be underestimated. A clinician would be required to provide support and individualized feedback and guidance, and to provide the link between the virtual environment and the real-world applicability. Technology can provide new interventions, or innovative delivery of approaches, but patients will continue to seek human clinicians who can empathise and help make sense of their experiences (Foley & Woollard, 2018). Particularly in ASD, it would be important to ensure the client did not become entirely dependent on the virtual environment (VE) and fail to apply their skills in the real world and interactions with real others. The clinician has the potential to provide this crucial link.

1.5. Neurodiversity and Implications for Intervention

One issue that is worthy of note is the criticism that some of the interventions outlined above have received, particularly those that are behavioural-based such as ABA, from families, autistic people and ethicists alike. The major issue is that target behaviours of intervention are often decided upon by non-autistic others, including therapists and parents (Milton, 2014). What may be thought of as problematic by parents and therapists, may not be considered so by the autistic person, who may have very different opinions on what they consider the most challenging, in terms of communication, social interaction, anxiety management or otherwise.

Many interventions can be criticised as approaching intervention from an entirely deficit-based model. Whilst there are several areas that individuals with autism find difficult, viewing the condition as entirely problem-based and implementing intervention as such has its disadvantages; it fails to consider the neurodiversity of human cognition, and more generally of the human condition. The concept of neurodiversity suggests that differences and variations in neurological development, neural functioning and related behavioural traits are a natural part of human diversity, and not to be pathologized under a medical model (Armstrong, 2011). Framing autism and other neurodevelopmental conditions as disorders to be ‘fixed’ implies a clinical ‘ideal’ of being human. An alternative approach under a neurodiversity framework advocates the implementation of support systems for independent living, assistive technologies, and training. This is not to underestimate the realities of individuals with high support needs, but rather to honour and maintain respect for the diversity of humanity and self-expression.

Interventions delivered with VR potentially face similar criticism, with intervention being developed from the perspective of researchers and clinicians. Whilst individuals with ASD are said to ‘lack theory of mind’, this lack can equally go in both directions as non-autistic people

often fail to consider the distress that might be caused by inhibiting what they deem negative or 'inappropriate' behaviours in ASD, which may in fact be used as methods of self-soothing or anxiety management.

It is imperative therefore that autistic individuals themselves are consulted about what they want from intervention, what they find the most challenging in day-to-day life, and what they might consider useful from intervention/training to empower them to live the life they want. Of course, this may be harder with children and those who may be less verbal or with lower cognitive abilities. However, intervention from a neurodiversity perspective would encourage therapists and allow individuals to engage with autists' interests, take into account autistic sensory differences and cognitive styles, and generally allow greater mutual respect in interventive interactions.

1.6 Use of Virtual Reality in Psychotherapeutic Intervention

An initial availability search of literature on the use of VR for individuals with and without ASD was conducted, and a brief overview of some of these different types of study are outlined in the sections that follow.

Researchers have begun to explore the utility of VR technology as a complementary tool in psychosocial intervention for individuals in the general population and in ASD (e.g. Halabi et al., 2017; Kandalaft, Didehbani, Krawczyk, Allen, & Chapman, 2013; Maskey, Lowry, Rodgers, McConachie, & Parr, 2014; Moore, Cheng, McGrath, & Fan, 2005; Newbutt, Sung, Kuo, & Leahy, 2016; Rothbaum et al., 1999; Rothbaum, Hodges, Watson, Kessler, & Opdyke, 1996).

1.6.1 Use of VR in non-ASD populations

Virtual reality has been used as an effective complementary psychotherapeutic intervention tool in treating various conditions in non-ASD populations. It has been used to treat anxiety and specific phobias such as fear of flying (Rothbaum, Hodges, Watson, Kessler, & Opdyke, 1996); body image and eating disorders (for a review see Ferrer-García & Gutiérrez-Maldonado, 2012). It has also been used in neurological conditions, such as stroke (Broeren, Claesson, Goude, Rydmark, & Sunnerhagen, 2008; Cameirão, Badia, Oller, & Verschure, 2010) to aid recovery, as well as post-traumatic stress disorder (Rothbaum et al., 1999). Specifically in relation to anxiety disorders, the use of VR has been proposed as a new complementary tool for exposure therapy (Rothbaum, Hodges, & Kooper, 1997). Virtual Reality appears to be a good alternative to, and

can offer a number of advantages over, *in vivo* exposure and imaginal exposure in that it can be provided in office-based traditional therapeutic settings, and is more controlled and cost-effective than *in vivo* exposure (Riva, 2005).

1.6.2 Use of VR for individuals with ASD

1.6.2.1 Benefits of VR for individuals with ASD

The benefits of VR technology in interventions for ASD individuals are numerous, and include:

- Providing a safe space in which to explore and experiment without unpleasant social consequences.
- Providing an intermediate step between imaginal exposure and a real-world scenario to enable emotion regulation techniques
- Offering an alternative to imaginal exposure, which may be difficult for ASD clients
- Controlling the intensity of a scene in preparation for the real world
- Circumventing time and financial limitations that might prevent a clinician from offering in vivo exposure
- Offering flexibility and individualised content for this heterogeneous population

Virtual reality provides a 'safer' environment for participants to practise and experiment without negative consequences, thus allowing the development of skills involved in activities of daily living (Strickland, 1998). Virtual reality environments have the potential to leverage some ASD strengths and traits. Individuals with ASD are very often engaged by computers and technology, and as such a VR environment has the potential to increase their engagement and motivation in an intervention offered in this modality (Parsons & Mitchell, 2002). Individuals with autism report that their thinking and learning is primarily visual in nature and there is evidence that ASD individuals respond better to visual methods of learning (O'Riordan, Plaisted, Driver, Baron-Cohen, 2001). The VR environment emphasises visual presentations and responses over other senses (Strickland, 1998) and as such has the potential to strengthen skill learning opportunities.

Use of VR may aid in keeping levels of anxiety within a manageable range to be able to interact with the therapist and the VR environment. Keeping anxiety levels within a workable range aids learning of further coping skills and social skills, that otherwise may not be possible when faced with real-world exposure which, for ASD individuals, can increase anxiety levels sharply (Maskey

et al., 2014). VR offers a unique intermediate step between imaginal exposure or basic visual exposure and real world interactive exposure. Given that individuals with ASD have been shown to be slower at self-regulating emotion and slower to return to baseline levels of arousal (e.g. Loveland, 2005), this intermediary step may be crucial in assisting ASD participants to develop the skills necessary to face the real-world situations which generate the most difficulties for the individual.

Sensory input of virtual environments can be controlled and made progressively more complex at a rate suitable for the individual (Strickland, 1998). For instance, more people can be added to a given scene, or noise levels can be increased, gradually ensuring the individual is comfortable with each level of increase. This allows individuals to gradually lower their anxiety levels and increase their confidence in the VE to then transfer to the real world (Strickland, 1998), and prepares the individual for the true nature of real world complex scenarios. Importantly, the individual can dictate the pace at which scenes become increasingly more complex.

Individuals with ASD have been reported to have limited imagination potential (Wing and Gould, 1979). As most exposure interventions for fear/phobia involve a step of imaginal exposure, the ASD client's limited capacity in this regard can limit techniques used in regular exposure therapy. VR thus provides the visual content that individuals cannot provide themselves with their imagination and an important step in making it generalisable to the real world.

Clinic time constraints and financial limitations of services do not always allow for real world exposure for anxiety reduction or for real world social skill building in different contexts. A lack of real world training may hinder the potential for generalisation to real life scenarios (Kandalaft et al., 2013). VR provides an opportunity to engage and interact in various scenes without in effect leaving the clinic room, making it cost and time effective, and additionally offers a step that is likely to increase transfer of strategies learned in the VRE to the real world.

VR offers control and importantly flexibility. ASD is a heterogeneous disorder and as such predefined intervention content cannot address the needs of all individuals with the condition. The flexibility that VR affords, allows individualisation of content and additional learning supports (Wang & Reid, 2011). This individualisation can be achieved by incorporating the client's specific stereotyped interests within the intervention to increase engagement and attention, and by managing the complexity of the displayed scenes at an appropriate rate for the client.

1.6.2.2 Use of VR as a learning aid in ASD

The type of virtual reality technology that has been used so far to address issues experienced by people with ASD has varied widely, from studies employing flat screen monitor 3D display environments (Cheng & Ye, 2010; Didehbani, Allen, Kandalaft, Krawczyk, & Chapman, 2016; Kandalaft et al., 2013; Mitchell, Parsons, & Leonard, 2007; Moore et al., 2005), 360 degree immersive environments (Maskey et al., 2014), head-mounted display technology (Newbutt et al., 2016), amongst others (Bauminger et al., 2007; Herrera et al., 2008).

Early studies have found that VR can be utilized by children with ASD as a learning tool to teach safety skills (Strickland, Mcallister, Coles, & Osborne, 2007; Josman, Ben-Chaim, Friedrich, & Weiss (2008), to hold their interest (Cobb et al., 2002), to monitor eye gaze in social interactions (Lahiri, Warren, & Sarkar, 2011), aid learning of pretend play (Herrera et al., 2008), and interpret emotions of avatars (Cheng & Ye, 2010).

1.6.2.3 Use of VR for psychosocial intervention in ASD

An initial availability search shows that VR for psychosocial purposes has been used with ASD individuals to develop social cognition skills with respect to learning social norms and social decision-making (Mitchell et al., 2007; Parsons, Leonard, & Mitchell, 2006), for emotion recognition (Didehbani et al., 2016; Kandalaft et al., 2013), to develop conversation skills (Kandalaft et al., 2013), to reduce anxiety (Austin, Abbott, & Carbis, 2008; Maskey et al., 2014), and to develop collaboration and reciprocity skills (Parsons, 2015).

1.6.2.4 Feasibility

One of the aims of this systematic review is to examine the feasibility of VR intervention in general for ASD individuals. However, some researchers have been inclined to ask whether specific VR technology is suitable for use with individuals with ASD, or whether there are VR equipment variables that would make it an unsuitable intervention aid, given some ASD traits such as sensory sensitivities. Newbutt, Sung, Kuo, & Leahy (2016) conducted a study to examine just this, specifically in relation to the use of head-mounted display (HMD) virtual reality equipment. Participants were recruited to two phases of the study: in the first, participants viewed three VR environments wearing HMDs (Oculus Rift technology) and headphones for a short period of time. In the second phase, exposure periods to VR environments were longer (25 mins) and more intense in terms of interactivity and environment. The researchers assessed

participants' overall enjoyment and sense of presence with the Sense of Presence Inventory. Participants' reported mean score following HMD-VR use on this measure was 4.32, (five = highest presence). Psychological Status was assessed with the State-Trait Anxiety Inventory before and after the VR experience. Participants mean score post-VR exposure was two (one = lowest levels of anxiety), and there was no anxiety change pre- and post- HMD-VR use. The study demonstrates that the use of VR presented in the form of HMD is not aversive and is a feasible tool for individuals with ASD, with the potential for a meaningful impact on how individuals develop skills to integrate into real life and overcome daily life challenges.

1.6.3. Reviews to date

With increased use of technology in recent years, more reviews are being published. Mishkind, Norr, Katz and Reger (2017) provided a review of VR intervention in psychiatry in general, but focussed on mental health disorders and not ASD. Liu, Wu, Zhao and Luo (2017) published a review on technology facilitated diagnosis and treatment in ASD. However, their review focusses on engineering elements of intervention components. Lau, Smit, Fleming and Riper (2017) conducted a systematic review of the use of serious games for intervention of mental health disorders. Although they included ASD participants, not all serious games are considered VR and not all VR applications are serious games. Mesa-Gresa, Gil-Gómez, Lozano Quilis and Gil-Gómez (2017) published a systematic review of VR use with ASD children. However, their review focussed entirely on the use of VR with young people up to the age of 18. Additionally, it included all general studies employing VR with children, and therefore did not examine the efficacy of VR as a potential intervention with pre-post treatment or change over time data.

1.7. The present thesis

1.7.1 Aims

To the best of our knowledge, no systematic review has been conducted of the current research relating to the use of VR in intervention for individuals (children and adults) with ASD, to address either anxiety or social communication and interaction. A full systematic review would be useful for researchers and clinicians alike in order to determine where research currently stands with respect to the use of VR technology for ASD intervention, and to determine the next steps in order to utilise VR most effectively and consider potential improvements in the current available interventions for ASD-related difficulties and anxiety.

More specifically, the present systematic review aims to examine the feasibility of VR use in interventions for ASD individuals (both children and adults); and aims to examine the utility of VR-based intervention for ASD individuals. To examine the utility of VR intervention, consideration will be given to efficacy (results under ideal controlled conditions); effectiveness of VR intervention (results under less than ideal natural clinical settings), and impact on real-world living for individuals.

2. METHODS

A search strategy was developed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA)-P guidelines (Moher et al., 2015), and the PRISMA 2009 Checklist (in Moher et al., 2009) to ensure the reporting quality of the systematic review.

2.1. Search Strategy

2.1.1. Database search

To ensure a comprehensive search of the subject area, we conducted an electronic search of the following databases: PsychInfo (1806 to 2018), Ovid Medline (1946 to 2018), Embase Classic + Embase (1947 to 2018) for papers published between 1990 and 2018 and in the English language. The last date for this search was the 10th December 2018.

2.1.2. Search terms

The database search used broad search terms and keywords to increase the likelihood of relevant studies being captured. The following search terms were used:

- "autism" OR "autism spectrum disorder" OR "autism spectrum condition" OR "ASD" OR "ASC" OR "Asperger*" OR "pervasive developmental disorder not otherwise specified" OR "PDD-NOS";
- AND "virtual reality" OR "virtual environment" OR "virtual reality environment" OR "VR" OR "VE" OR "3D" OR "3-Dimens*" OR "Virtual Display" OR "Head-mounted Display" OR " HMD".

2.2. Inclusion and Exclusion Criteria

2.2.1. Focus of Study

The papers considered for inclusion in the systematic review reported use of virtual reality as part of intervention with individuals with ASD for either (i) core deficits in social communication and interaction, or (ii) general anxiety or specific anxiety disorders. The reason for examining these two areas is due to the substantial inter-relatedness of anxiety and core deficits of ASD and how they impact on each other (MacNeil et al., 2009; Simonoff et al., 2008; White et al., 2014)). The review did not consider for inclusion studies which aimed to examine change in

functioning in non-ASD related areas (e.g. development of maths skills), unless secondary outcome measures were reported relating to ASD traits and/or anxiety.

2.2.2. Study participants

The systematic review considered studies published with participant ASD children, adolescents and adults. Studies considered for inclusion reported participants to have a confirmed diagnosis of autism, autism spectrum disorder, Asperger's Syndrome or PPD-NOS. Studies were excluded if the recipient participants of the intervention/ training were tutors of individuals with ASD receiving training in interventions with VR.

2.2.3. Type of study

An initial availability search identified a limited number of studies utilizing VR as an intervention in autism. For this reason, all study designs were considered for inclusion in the systematic review; that is, not only randomized controlled trials, but single-arm studies, group-based experimental designs and case studies. Pagliaro, Bruzzi, and Bobbio (2010) recognise the quality of randomised controlled trials, but suggest they are not as widely used in clinical decision making due to their emphasis on methodology and rigour rather than on clinical relevance. This systematic review aims to collate evidence in respect of intervention with VR irrespective of design at this stage and due to prematurity of the area.

2.2.4. Type of intervention

The review considered interventions in all forms, including social skills training and psychosocial or psychotherapeutic interventions in any modality (e.g. cognitive behavioural therapy, acceptance and commitment therapy).

2.2.5. Type of VR technology

Whilst the implementation of head-mounted display and 360 VR is increasing, the use of this type of VR technology with individuals with autism is at present limited. Therefore, the review considered the use of VR in its broadest interpretation (i.e. interactions with 3D digitally generated environments; Mon-Williams & Wann, 1996), which includes head-mounted display devices, 360 degree (CAVE) environments, half-CAVE environments, 3D flat-screen displays with avatars (such as Second Life; Linden Lab, 2003), and touchscreen monitors with 3D graphics. The review did not consider interventions with flat screen monitor 2D displays.

2.2.6. Outcome measures

The review considered publications that reported pre- and post-intervention outcome measure data, or reported change over time with intervention sessions, relating to social communication and interaction and/or anxiety measures. More specifically, the review considered studies assessing social skills abilities (verbal and non-verbal), social cognition, social confidence, or anxiety levels. Studies included considered outcomes as assessed by the intervention provider or independent evaluators, or the participants' own outcome evaluation.

2.2.7. Exclusion Criteria

- Studies that do not provide pre- and post-intervention/training data or change over time with intervention sessions.
- Studies that use only flat-screen monitor 2D displays.
- Studies whose participant recipients of the intervention/training are not individuals with ASD.

2.3. Study selection and screening

The following process for study selection and eligibility screening was conducted:

1) The systematic search was conducted by the author (VB) according to the strategy described above. All citations identified were transferred to Excel and duplicates removed by digital object identifier (DOI) reference.
2) In the first stage of screening, the identified titles and abstracts were screened by the author (VB) for inclusion eligibility. If an abstract was unavailable or did not contain adequate information, the article was submitted to the second phase of selection screening for full article review. Articles that were considered ineligible for inclusion were excluded and the reason coded.
3) In the second eligibility phase, articles meeting this stage were full-text reviewed by the author and assessed for meeting inclusion criteria. Articles that were considered ineligible for inclusion were excluded and the reason coded.
4) A random selection of 20% of articles identified in the second stage were reviewed independently by another reviewer (ES, a research assistant). Differences of opinion between the two reviewers (VCB and ES) on the selection of studies were resolved by

consensus with the supervising author (FM). If consensus could not be reached, the protocol indicated that the supervisor's decision would be considered final.

5) Of the final selection of papers identified as meeting inclusion criteria for the review, full-text articles were double checked / cross referenced to identify any duplication of reported data across articles.
6) Data relevant to the review were extracted from each included article and inputted into Excel.

2.4 Data extraction

Once the final included literature was decided upon from the full text review, data extraction was completed by the author. Extraction was performed in Excel with a combination of tick boxes and text. The following parameters of the included literature were recorded:

- Study characteristics
  - authors
  - year of publication
  - journal name
  - location of study
  - reference
  - type of study
- Participant characteristics
  - number of participants
  - diagnosis
  - age range of participants
  - inclusion and exclusion participant criteria
  - comparator participant group (if used)
- Aims of study
- Intervention method
  - intervention approach (e.g. ABA, CBT, ACT)
  - type of VR technology used
  - VR environment(s) employed
  - VR intervention scenario / task aim
  - number of intervention sessions
  - duration of intervention sessions

  - intervention components
  - comparator interventions (if used)
- Outcome measures used to assess competences in social communication, social interaction, or social cognition, and outcome measures of anxiety ratings. To be repeated for each outcome measure:
  - Outcome measure name
  - Outcome definition
  - Time points measured
  - Unit of measurement
  - Person measuring/reporting
  - Is outcome tool validated?
  - Imputation of missing data
  - Follow-up assessed?
  - Follow-up in real word assessed?

2.5 Quality Assessment

An assessment of the quality (or risk of bias) of studies included in a review is considered an important component of a systematic review, and contributes to the evaluation of the strength of evidence presented (Shamseer et al., 2015). Different organisations have been established dedicated to assessing quality in reviews (e.g. Agency for Healthcare Research and Quality (AHRQ), The Cochrane Collaboration, The NHS Centre for Reviews and Collaboration, The York Centre for Reviews and Dissemination) and methodologies for assessing quality of studies have been documented (Higgins & Green, 2011; Viswanathan, Ansari, Berkman, Chang, Hartling, McPheeters, Santaguida, Shamliyan, Singh, Tsertsvadze, 2012).

Several quality assessment tools for individual studies have now been published (Zeng et al., 2015). The aim of such tools is to seek objectivity in the way research evidence is judged. The large majority of these tools, however, have been developed to assess randomised-controlled trials (RCTs). However, for some areas of healthcare, few RCTs exist or conducting them is not considered feasible (Reeves, Deeks, Higgins, Wells, 2011) . Furthermore, premature areas of research, such as that which this present systematic review examines, are possibly not at a stage for an RCT to be conducted. Nevertheless, it is equally important for such research to be assessed for bias. Indeed, potential biases can be greater for non-randomised studies compared with randomised trials, particularly with respect to differences between people in different

intervention groups (selection bias) and studies that do not explicitly report having had a protocol (reporting bias) (Reeves et al.,2011).

Increasingly more tools have been proposed to assess the quality of non-randomised studies studies (Deeks et al., 2003; Sanderson, Tatt, & Higgins, 2007). The Newcastle-Ottawa (Wells, Shea, O'Connell, Peterson, Welch, Losos, Tugwell, 2008) and the Downs and Black (Downs & Black, 1998) tools were both shortlisted for being methodologically sound (Deeks et al., 2003). These two and the more recent ROBINS-I tool (Sterne et al., 2016) were initially considered for use in the quality assessment of included studies in this present systematic review. The Robins-I tool and the Newcastle-Ottawa scale both provide means to assess non-randomised research studies only, whilst the Downs and Black tool allows for the quality assessment of both randomised and non-randomised studies and provides reliability and validity data on the assessment tool for randomised and non-randomised studies separately as well as combined. As the present systematic review yields studies that involve randomised as well as non-randomised methods, the Downs and Black tool was selected for the risk of bias assessment of included studies (provided in Appendix B).

Irrespective of the study design type (e.g. randomised-controlled trials, case-control studies, single-arm non-controlled studies), all study designs aim to test an association between intervention and the outcome, and aim to minimise flaws that may bias the association (confounders). The Downs and Black assessment tool evaluates individual studies' ability to do this through 27 item questions in the areas of five principal constructs: (i) Reporting (10 items), addresses whether the information provided in a paper is sufficient and clear enough to allow a reader to make an unbiased assessment of the findings of the study; (ii) External validity (3 items), addresses the extent to which findings can be generalised to the population from which the study participants are derived; (iii) Internal validity - Bias (7 items), addresses biases in the measurement of the intervention and the outcome; (iv) Internal Validity - Confounding (6 items), addresses bias in the selection of study participants; and (v) Power (1 item), assesses whether negative findings from a study could be due to chance.

The Downs and Black tool demonstrated high internal consistency for the Quality Index (KR:20 = .89) for both randomised and non-randomised studies. The subscales showed adequate internal consistency, with the exception of the External Validity subscale. The Quality Index also demonstrated good test-retest reliability (r = .88), as well as good inter-rater reliability r = .75), and good criterion validity (r = .98) (Downs & Black, 1998).

One assessor (principal author, not blinded to study) conducted the quality assessment for each of the included studies in the systematic review. A second assessor conducted the quality assessment on a selection of the included studies (three papers of different study type) to ensure agreement on how the Downs and Black tool were to be used across study design.

For each of the 27 items, a score of 0 or 1 is given, except for one item within the Reporting subscale, which is scored 0 to 2, and a single item on the power sub-scale that is scored 0 to 5. The maximum total score a study can achieve is 32. Some questions of the Quality Assessment were not deemed applicable for non-controlled studies and case studies. As such, these questions were noted as Not Applicable (NA) for these studies and the questions were not counted towards the final possible total score that study could receive. A final percentage score was given to each study based on total score received out of total applicable score.

It should be noted that as a different number of question items contribute to each of the subscales of the Downs and Black Quality Assessment that the subscales thus contribute different weightings to the overall assessment score. Each subscale, however, may be of more or less significance depending on the type of study (e.g. selection and reporting bias may be more important for non-RCTs, Reeves et al., 2011). The Quality Assessment does not take this into account or allow for different weightings across studies. Whilst this may be considered a flaw of the Quality Assessment, the use of a percentage score of total applicable items nevertheless allows an evaluation of the studies relative to one another. Arbitrary threshold values were determined to loosely assign categories of 'low/poor' quality (<50% of total applicable score), 'moderate' (50-68%), and 'good' quality (>68%). Commentaries on the quality of studies are made relative to other included studies in the review. The Downs and Black quality assessment tool will not be used as a means to include or exclude papers from the review, but rather as means to evaluate the quality of those included according to the selection criteria provided above (Section 2.2). The Quality Assessment tool will be utilised to highlight those studies deemed as higher quality and the results yielded from these, and consider next steps forward with these in mind in the Discussion.

2.6 Narrative synthesis

A systematic narrative synthesis will be provided, with information gathered from the data extraction presented in tables and text in order to summarise characteristics and findings of the included studies. The narrative synthesis will also explore the findings and relationships within

and between included studies, in line with the guidance on the conduct of narrative synthesis in systematic reviews from the Economic and Social Research Council (Popay et al., 2006).

3. RESULTS

Figure 1 shows the process of screening in a PRISMA flow diagram format. 1330 articles were returned in the initial systematic search. 610 duplicates were identified by Digital Object Identifier (DOI), with 720 papers remaining. After title and abstract screening, 116 articles were submitted to full text review. Of these articles, four papers were found to report duplicate data findings from other papers from the same lab group. A total of 24 articles were considered eligible for the review.

Table 1 summarises the extracted data with respect to study characteristics and participant characteristics of the 24 included papers in the review.

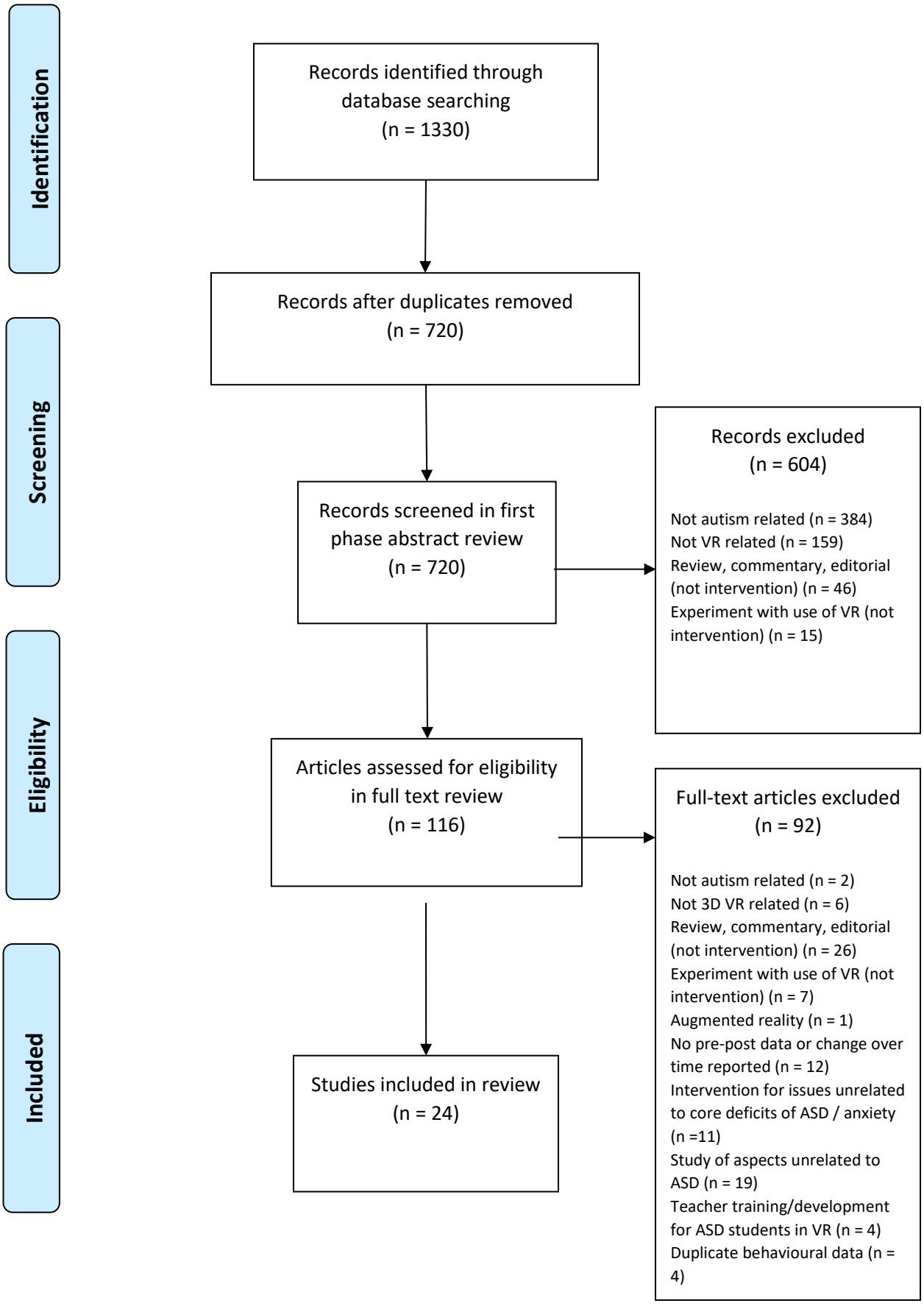

Figure 1. PRISMA flow diagram of review results.

Table 1. Table of study and participant characteristics

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Amaral et al., 2018	Frontiers in Neuroscience	Portugal, University of Coimbra	Single arm pre-post-intervention. 6 month follow-up. clinical feasibility trial. Brain-Computer Interface (EEG) for social cognition skills training.	15 (15 male)	ASD (by ADOS and ADI-R and/or DSM-V) as assessed by a psychologist and paediatrician	16 to 38	Excluded if: FSIQ < 80, had a neurological condition (e.g. epilepsy), other genetic syndrome, or other diagnosed comorbidities.	None
Burke, Bresnahan, Li, Epnere, Rizzo, Partin, Ahlness & Trimmer, 2018	Journal of Autism and Developmental Disorders	USA Florida International University	Single arm study. Pre-post and multiple time probe, comparison of interview skills following ViTA system job interviewing skills practice	32 (25 male)	ASD (22), intellectual disability (11), reported other disabilities (8) confirmed by vocational rehabilitation record or psychology report. Not specified how many ppts had both ASD and intellectual disability.	19 to 31 years (mean = 23, SD = 3.12)	Not specified	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Cheng, Chiang, Ye & Cheng, 2010	Computers and Education	Taiwan, China	Single arm, multiple baseline, multiple-probe (baseline, intervention, maintenance) study	3 (3 male)	Reported as having impaired empathy or symptoms consistent with ASD according to school records.	8 to 10 years	Included if: verbal IQ, performance IQ, and full-scale IQ >70 on WASI-III.	None
Cheng & Huang, 2012	Research in Developmental Disabilities	Taiwan, Republic of China	Single subject, multiple-probe (baseline, intervention, maintenance) multiple-baseline, across subjects design.	3 (3 male)	PDD with deficits in joint attention skills (according to parental records)	9 to 12 years	Pre-requisites of WASI VIQ, PIQ and FSIQ all > 50	None
Cheng, Huang & Yang, 2015	Focus on Autism and Other Developmental Disabilities	Taiwan, Republic of China	Single subject, single arm, multiple-probe across ppts (baseline-intervention-maintenance) design.	3 (3 male)	ASD according to school reports of social impairments consistent with ASD, FSIQ>80.	10-12 years	Not specified	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Cheng, Luo, Lin & Yang, 2018	International Journal of Developmental Disabilities	Taiwan, Republic of China	Pre- post-test, randomised control vs treatment with 3D Complex Facial Expression Recognition (3CFER) system	24 (12 treatment (8 male), 12 control (8 male))	ASD	Treatment group = 9.2 to 12.7; Control = 9.5 to 12.6	Included if WASI-IV VIQ, PIQ and FSIQ > 70. Excluded if presence of an emotional disorder as assessed by Scale for Assessing for Emotional Disturbance (SAED, Epstein & Cullinan, 1998)	Use of paper based emotion pictures and social stories for recognising emotions (Howlin et al., 1999)
Cheng & Ye, 2010	Computers and Education	Taiwan, Republic of China	Single arm, multiple baseline, multiple probe (baseline, intervention, maintenance) study	3 (2 male)	Reported as having impairments in social competence consistent with autism spectrum disorder according to their school record	7-8 years old	Included if: verbal IQ, performance IQ, and full-scale IQ >70 on WASI-III.	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Didehbani, Allen, Kandalaft, Krawczyk & Chapman,2016	Computers in Human Behaviour	Dallas Texas, USA, University of Texas	Single arm pre-post social skill training study	30 (26 male)	Asperger's or PDD (confirmed for research purposes by trained clinicians with the ADOS)	7 to 16 years	Excluded if had acute psychiatric condition or Axis I psychopathology except managed ADHD. Thirteen ppts had comorbid ADHD. All had estimated IQ in average to above-average range on WASI.	None
Herrera, Alcantud, Jordan, Blanquer, Labajo & De Pablo,2008	Autism	Valencia, Spain	Case studies. Single arm pre-post design.	2 (2 male)	Formal ASD diagnosis with DSM-IV	8:6 and 15:7	None reported	None
Ip, Wong, Chan, Byrne, Li, Yuan, Lau & Wong, 2018	Computers and Education	Hong Kong	Pre- post intervention. Training vs Control comparison study using VR based intervention to enhance emotional and social adaptation	94 children recruited (47 in each group - 42 male in Training group, 44 male in control). 36 in each group included in analysis (31 male in treatment	Clinical diagnosis of ASD (not confirmed)	6 to 12 years old. But 7 to 10 years included in analyses. Training group M= 107.6 months (SD = 13.27); Control group M = 104.8 (SD =	None specified. Ppts included in final analyses were filtered for age (7 to 10 years)	Waitlist control (no training)

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
			skills in children	group, 33 in control).		13.83)		
Kandalaft, Didehbani, Krawczyk, Allen, Chapman, 2013	Journal of Autism and Developmental Disorders	Texas, Dallas, US	Single arm pre-post social skill training study. 6 month follow-up survey.	8 (6 male)	Asperger's or PDD from psychiatrist, confirmed by ADOS	18-26	Excluded if had acute psychiatric condition or Axis I psychopathology (except managed depression and history of neurological disorders)	None
Ke, Im, 2013	The Journal of Educational Research	Florida USA. Sessions held at location of child's preference (home, parents' offices or school)	Pre- post intervention. Multiple-baseline across ppts design. Social skill training.	4 (2 male)	Formal medical or educational diagnosis of Asperger's Syndrome.	9 to 10 years	Included if verbal. Exclusion criteria not defined.	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Ke, Lee, 2016	Interactive Learning Environments	Florida USA.	Qualitative single subject case studies, ASD vs TD ppts. Qualitative time-series and micro behaviour analysis.	2 ASD (1 male), 1 TD (female)	Formal medical diagnosis of Asperger's Syndrome	ASD: aged 9 and 10; TD: aged 8	None specified	1 TD child
Lorenzo, Lled, Pomares, Roig, 2016	Computers and Education	Alicante, Spain	Experimental study - Immersive VR vs flat-screen VR for improving emotion skills in children with ASD	40 in total. 20 in experimental group, (14 male) 20 in control (15 male).	ASD	7 to 12 years	Pupils selected following interview confirming difficulties typical of ASD	VR software application. Comparison group = intervention with computer screen VR.
Lorenzo, Pomares, Lledo, 2013	Computers and Education	Alicante, Spain	Single arm, mixed methodology: qualitative, observational, experimental and quantitative. Comparison between first and last intervention session.	20 (16 male)	Asperger's	8 to 11 years (primary school), and 12 to 15 years (secondary school)	None specified	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Maskey, Lowry, Rodgers, McConachie, Parr, 2014	PLOSOne	Newcastle, UK.	Single-arm pre-post intervention comparison.	9 (9 male)	ASD, corroborated by MDT clinicians. No reported learning disability. SCQ scores were compatible with ASD diagnosis.	7 to 13 years	None specified	None
Mitchell, Parsons, Leonard, 2007	Journal of Autism and Developmental Disorders	Nottingham, UK	Single arm pre-post-intervention study. Two pre-intervention baseline periods between participants	6 (3 male)	Formal diagnosis of Autism, Asperger, ASD using DSM-IV (not confirmed for research), reported FSIQ range = 65 to 110	14 to 16 years	Included if no comorbid diagnoses	None
Parsons, Leonard, Mitchell, 2006	Computers and Education	Nottingham, UK	Qualitative case studies.	2 (males)	Special Educational Needs school statement: "demonstrating behaviours associated or consistent with autism including	14, 17	Selected as having 'good' verbal skills. VIQ, PIQ, FSQ all >70	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
					difficulties with social understanding"			
Serret, Hun, Iakimova, Lozada, Anastassova, Santos, Vesperini, Askenazy, 2014	Molecular Autism	France	Single arm pre-post training study.	33 (31 male)	ASD (high and low functioning) (Autism (23), Asperger (4), PDD-NOS (6), WASI FSIQ range = 35 to 129, M = 70.05 (27.6)). ASD determined according to DSM-IV with ADI-R and/ADOS	6 to 17 (M = 11.4, SD = 3.16)	Included if: able to discriminate primary and secondary colours, had already used a computer	None

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
Smith, Ginger, Wright, Wright, Taylor, Humm, Olsen, Bell Fleming, 2014	Journal of Autism and Developmental Disorders	Chicago, USA, Northwestern University	Single-blinded randomised control trial. Pre-post training with Virtual Reality Job Interview Training (VR-JIT).	26 in total. 16 intervention group (12 male), 10 Control (8 male)	ASD. Confirmed with parent and self-report Social Responsiveness Scale, 2nd edition t-scores >60	18 to 31 years. TAU mean = 23.2 (3.0); VR-JIT mean = 24.9 (6.7)	Inc if: 6th grade reading level determined by Wide Range Achievement test-IV (Wilkinson & Robertson 2006), willingness to be video-recorded, working less than part-time, actively seeking employment. Exc if: medical issue compromising cognition (e.g. TBI), uncorrected vision or hearing problems, diagnosis of substance abuse/dependence	Treatment as Usual
Smith, Fleming, Wright, Losh, Humm, Olsen, Bell, 2015	Journal of Autism and Developmental Disorders	Chicago, USA, Northwestern University	6 month follow-up survey study to examine vocational outcomes following RCT VR job interviewing	23 (out of the 26 ppts from study above Smith et al., 2014) (15 - VR-JIT treatment group, 8 Control)	ASD. Confirmed with parent and self-report Social Responsiveness Scale, 2nd edition t-scores >60	18 to 31 years. TAU mean = 23.2 (3.0), VR-JIT mean = 24.9 (6.7)	Inc if: 6th grade reading level determined by Wide Range Achievement test-IV (Wilkinson & Robertson 2006), willingness to be	Treatment as Usual

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
			training vs treatment as usual.				video-recorded, working less than part-time, actively seeking employment. Exc if: medical issue compromising cognition (e.g.TBI), uncorrected vision or hearing problems, diagnosis of substance abuse/dependence	
Stichter, Laffey, Galyen, Herzog, 2014	Journal of Autism and Developmental Disorders	Missouri, USA	Single arm pre-post intervention group design. A pilot field test.	11 (gender not reported)	ASD (confirmed with ADOS and ADI-R)	11 to 14 years (M = 12.57, SD = 0.75)	Clinical diagnosis of ASD, FSIQ > 75, access to TD peers for part of their day	None
Strickland, Coles, Southern, 2013	Journal of Autism and Developmental Disorders	Atlanta, Georgia, USA	Randomised treatment vs control. Pre- post training comparison.	22 in total. 11 in treatment group, 11 in control (all male)	High functioning autism or Asperger's Syndrome	16 to 19	To have regular access to a home computer with internet, could perform basic computer and website navigation functions indepen-dently. Excluded if: vision, hearing, or	No treatment

Authors & Publication Year	Journal	Location of study	Study type	No of participants	Diagnosis	Age range in Years	Participant inclusion/ exclusion criteria	Comparison group
							motor problems that prevent participation in VR practice session or interview simulations, never been competitively employed nor viewed the JobTIPS website.	
Yang, Allen, Abdullahi, Pelphrey, Volkmar, Chapman, 2017	Behaviour Research and Therapy	USA. Child Study Centre, Yale University School of Medicine	Single arm, pre-post-treatment only clinical trial. Virtual Reality Social Cognition Training. fMRI neuroimaging study.	17 (15 male)	ASD, according to DSM-V and using ADOS (re-confirmed) as assessed by research-reliable clinicians and clinical psychologists	18.06 - 31.08. Mean = 22.5, SD =3.85	Inc if: FSIQ > 80, diagnosis of ASD, 18-40 years old, mean length of utterance > 5 words, free of metal implants, no evidence of claustrophobia. Exc if: Not fluent in English, significant hearing loss, serious head trauma, other psychiatric disorder.	None

3.1. Study and participant characteristics

3.1.1. Geographical location of studies

Of the total 24 papers included in the systematic review, eight studies were conducted in Europe: one in France (Serret et al., 2014), one in Portugal (Amaral et al., 2018), three in Spain (Herrera et al., 2008; Lorenzo, Lledó, Pomares, & Roig, 2016; Lorenzo, Pomares, & Lledó, 2013), and three in the United Kingdom (Maskey et al., 2014; Mitchell et al., 2007; Parsons et al., 2006). Nine studies were conducted in the United States (Burke et al., 2018; Didehbani et al., 2016; Kandalaft et al., 2013; Ke & Im, 2013; Ke & Lee, 2016; Smith et al., 2015, 2014; Stichter, Laffey, Galyen, & Herzog, 2014; Strickland, Coles, & Southern, 2013; Yang et al., 2017); and six studies in the Republic of China (Cheng, Chiang, Ye, & Cheng, 2010; Cheng, Huang, & Yang, 2015; Cheng, Luo, Lin, & Yang, 2018; Cheng & Huang, 2012; Cheng & Ye, 2010) (one in Hong Kong (Ip et al., 2018)).

3.1.2. Study type

Of the total 24 studies included in the review, 14 studies were single-arm studies with ASD participants only using either a pre-post intervention design (nine) (Amaral et al., 2018; Didehbani et al., 2016; Kandalaft et al., 2013; Ke & Im, 2013; Lorenzo et al., 2013; Mitchell et al., 2007; Serret et al., 2014; Stichter et al., 2014; Yang et al., 2017), or a multiple-probe design (i.e. baseline, intervention, maintenance) (five) (Burke et al., 2018; Cheng et al., 2010, 2015; Cheng & Huang, 2012; Cheng & Ye, 2010). One study compared effects of two types of VR intervention (i.e. two types of VR technology) with ASD participants (Lorenzo et al., 2016). Five studies employed a randomized controlled design (three with waitlist controls (Cheng et al., 2018; Ip et al., 2018; Strickland et al., 2013), two with treatment as usual (Smith et al., 2015, 2014)) . Three studies employed a case study design (Herrera et al., 2008; Ke & Lee, 2016; Parsons et al., 2006). Of these latter studies, one examined intervention effects comparing typically developing children with ASD children (Ke & Lee, 2016).

3.1.3. Size of study

The size of studies varied, with numbers of participants in the treatment/experimental group ranging from 2 to 33 participants. Seven studies had just two or three participants (Cheng et al., 2010, 2015; Cheng & Huang, 2012; Cheng & Ye, 2010; Herrera et al., 2008; Ke & Lee, 2016; Parsons et al., 2006); four studies had between four and ten participants (Kandalaft et al., 2013; Ke & Im, 2013; Maskey et al., 2014; Mitchell et al., 2007); nine studies had between 11 and 20 participants (Amaral et al., 2018; Cheng et al., 2018; Lorenzo et al., 2016, 2013; Smith et al.,

2015, 2014; Stichter et al., 2014; Strickland et al., 2013; Yang et al., 2017), and four studies had more than 20 participants in the VR intervention/experimental group (Burke et al., 2018; Didehbani et al., 2016; Ip et al., 2018; Serret et al., 2014).

3.1.4. Participant diagnoses

Fourteen studies reported participants as diagnosed with ASD (Amaral et al., 2018; Burke et al., 2018; Cheng et al., 2010, 2015, 2018; Cheng & Ye, 2010; Herrera et al., 2008; Ip et al., 2018; Lorenzo et al., 2016; Maskey et al., 2014; Smith et al., 2015, 2014; Stichter et al., 2014; Yang et al., 2017); one study reported participants diagnosed with autism (Parsons et al., 2006); four Asperger's (Ke & Im, 2013; Ke & Lee, 2016; Lorenzo et al., 2013; Strickland et al., 2013); one study with PDD (Cheng & Huang, 2012); two with either Asperger's or PDD (Didehbani et al., 2016; Kandalaft et al., 2013); and two studies reported including participants with varied diagnoses of autism, Asperger's, ASD or PDD, i.e. high and low functioning autism (Mitchell et al., 2007; Serret et al., 2014). One study reported participants as diagnosed with ASD with comorbid learning disability (Burke et al., 2018). Seven of the included studies re-confirmed the diagnoses for research purposes (Didehbani et al., 2016; Kandalaft et al., 2013; Maskey et al., 2014; Smith et al., 2015, 2014; Stichter et al., 2014; Yang et al., 2017).

3.1.5. Age of participants

The age of participants varied across studies. The majority of studies involved children and adolescents (up to the age of 18). Eighteen studies in total involved children and adolescents. Of these, ten studies involved primary aged school children (up to and including the age of 12) (Cheng et al., 2010, 2015, 2018; Cheng & Huang, 2012; Cheng & Ye, 2010; Ip et al., 2018; Ke & Im, 2013; Ke & Lee, 2016; Lorenzo et al., 2016, 2013); two studies involved adolescents (Mitchell et al., 2007; Parsons et al., 2006); and six studies involved both primary and secondary school aged children and adolescents (Didehbani et al., 2016; Herrera et al., 2008; Lorenzo et al., 2013; Maskey et al., 2014; Serret et al., 2014; Stichter et al., 2014). Five studies in total involved only adult participants (aged 18 and over) (Burke et al., 2018; Kandalaft et al., 2013; Smith et al., 2015, 2014; Yang et al., 2017), and two studies involved a mix of both children and adults (Amaral et al., 2018; Strickland et al., 2013).

3.2. Intervention Characteristics

Table 2 summarises the extracted data with respect to intervention characteristics of the 24 included papers in the review.

3.2.1. Focus of study

The studies included in the review varied in terms of the focus of the intervention. The majority of studies employed VR as a means to develop social communication (verbal and non-verbal) and social interaction skills (Amaral et al., 2018; Cheng et al., 2010, 2015; Cheng & Huang, 2012; Cheng et al., 2018; Cheng & Ye, 2010; Didehbani et al., 2016; Ip et al., 2018; Kandalaft et al., 2013; Ke & Im, 2013; Ke & Lee, 2016; Lorenzo et al., 2016, 2013; Mitchell et al., 2007; Parsons et al., 2006; Serret et al., 2014; Stichter et al., 2014; Yang et al., 2017). Two studies focussed specifically on improvement in non-verbal communication skills of joint attention (Amaral et al., 2018; Cheng & Huang, 2012). Five studies examined change in emotion recognition skills (Cheng et al., 2018; Didehbani et al., 2016; Lorenzo et al., 2016; Serret et al., 2014; Yang et al., 2017); one study examined change in (facial) emotional expression (Lorenzo et al., 2016), and one study in understanding and expression of empathy (Cheng et al., 2010). Three studies focussed on socio-cognitive reasoning and social decision-making (Mitchell et al., 2007; Parsons et al., 2006; Yang et al., 2017).

Four studies focussed specifically on increasing competency in job interviewing skills and increasing confidence levels in this context (Burke et al., 2018; Smith et al., 2015, 2014; Strickland et al., 2013).

Two studies examined general executive functioning (Didehbani et al., 2016; Lorenzo et al., 2013); one study examined improvement in symbolic play (Herrera et al., 2008). Finally, one study used VR intervention to reduce anxiety (Maskey et al., 2014).

3.2.2. Intervention approach

The intervention approach employed across studies also varied considerably, and some studies combined more than one approach. Two studies employed CBT or aspects of CBT (Maskey et al., 2014; Stichter et al., 2014). Two studies employed Social Stories (Cheng & Ye, 2010; Lorenzo et al., 2016, 2013). Seven involved facilitated practice of social situations with role-play and reward reinforcement (Ip et al., 2018; Ke & Im, 2013; Ke & Lee, 2016; Lorenzo et al., 2016, 2013; Parsons et al., 2006; Strickland et al., 2013). Three studies employed Social Cognition Training which also involved role-play practice of social situations (Didehbani et al., 2016; Kandalaft et al., 2013; Yang et al., 2017). Five studies explicitly reported that intervention involved hierarchical 'scaffolded' experiential learning (Burke et al., 2018; Cheng et al., 2015; Mitchell et al., 2007; Smith et al., 2015, 2014). Two studies were experimental and rewarded correct responses (Amaral et al., 2018; Serret et al., 2014). One study employed aspects of Applied Behavioural Analysis (ABA) (Stichter et al., 2014). Two studies employed video-modelling

(Herrera et al., 2008; Strickland et al., 2013). Two studies employed emotion control techniques and relaxation (Ip et al., 2018; Maskey et al., 2014).

3.2.3. VR technology type

Of the 24 studies included in the systematic review, 21 in total used a flat screen monitor of some kind. Twelve of these used a desktop computer with keyboard and mouse (Burke et al., 2018; Cheng & Huang, 2012; Didehbani et al., 2016; Kandalaft et al., 2013; Ke & Im, 2013; Ke & Lee, 2016; Serret et al., 2014; Smith et al., 2015, 2014; Stichter et al., 2014; Strickland et al., 2013; Yang et al., 2017); one used a flatscreen touchscreen monitor (Herrera et al., 2008), four used a laptop (Cheng et al., 2010; Cheng & Ye, 2010; Mitchell et al., 2007; Sarah Parsons et al., 2006); one used a tablet (Cheng et al., 2018). Four employed SecondLife software (Didehbani et al., 2016; Kandalaft et al., 2013; Ke & Im, 2013; Ke & Lee, 2016). Three studies employed head-mounted displays (HMDs): one Oculus Rift (Amaral et al., 2018), and one I-Glasses PC 3D Pro (Cheng et al., 2015). Two studies employed a four-sided CAVE environment (Ip et al., 2018; Maskey et al., 2014), and two a semi-CAVE (L-shaped) environment (Lorenzo et al., 2016, 2013).

Table 2. Intervention Characteristics

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Amaral, et al., 2018	(i) to assess feasibility and potential use of EEG-based BCI technology for ASD, (ii) assess use of neurophysiological rehabilitation tools for improving joint attention behaviour	Oculus Rift Development Kit-2 Head mounted display, with embedded eye tracking. 8-channel EEG BCI device	Joint Attention Assessment Task (JAAT) - Observation of avatar non-verbal social cues. Ppt identifies objects of interest. Direct feedback and reward.	JAAT task: Café, Classroom, Kiosk, Zebra Crossing. BCI task: bedroom with common furniture and objects	Avatar turns head or points at object of interest in scene. Ppt's attentional focus determined by eye tracking and EEG acquisition (oddball paradigm).	7 sessions over 4 months. First 4 sessions weekly, the remaining monthly.	Not specified	Viewing joint attention animations - avatar looks towards or points at object of interest. Animations repeated across settings. Number of items of appropriate social attention recorded and defined as eye fixations inside an area of interest after the start of the joint attention animation. Immediate feedback about attention given to ppt by target object flashing green if attended to, or red if not.	Control animations (no social skills relayed. e.g. avatar coughing, rolling or scratching head, yawning)

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Burke et al., 2018	To examine changes in competence and anxiety over time.	60-70 inch HD flatscreen monitor 6 camera system including XBOX GEN 1 KINECT camera to track facial expressions	Virtual Interactive Training Agent (ViTA) - a job interview role-play system, with hierarchical experiential learning. Explicit instruction, guidance and feedback.	Seven backgrounds: - hotel lobby, business office, warehouse breakroom (others not reported) to align with ppts' employment interests	Structured interview interactions ask ppts to (a) engage in social mores and introductory statements; (b) emphasise their strengths and have self advocacy; (c) self-promote; (d) provide a situational or behavioural example; (e) focus on practical housekeeping of the job opportunity; (f) recognise the job interview is coming to an end.	5 stages: 4 ViTA sessions, 1 face-to-face interview	Not reported	6 virtual humans with different behavioural dispositions. 10 to 12 interview questions in each session. Training opportunities provide interactions that progress in difficulty and are adjusted to needs of user. Ppts have coursework to learn core interviewing and employment skills. Explicit instructional strategies to respond appropriately on: interview etiquette (greetings, acceptable small talk, closing and thanking interviewer); how to make a good impression, provide clear and concise responses, identify individual strengths and how to self-promote, engage in active listening, convey interest using verbal and non-verbal communication.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Cheng, Chiang, Ye, Cheng, 2010	Investigate whether the Collaborative Virtual Learning Environment (CVLE) -3D Empathy System increases understanding of empathy (kindness, toleration, respect) expression of empathy.	Laptop, Windows XP, 3D environment developed using 3D Max, Virtools and Poser.	Collaborative Virtual Learning Environment CVLE -3D Empathy System - Facilitated practice of social situations. Reward and feedback on responses. Avatars to represent themselves.	Restaurant	Four social scenes: Someone cuts in the restaurant queue; a stranger sits next to the participant whilst he is eating; a drink is spilled on the floor; a passerby slips and falls. Ppts required to establish empathic situations and use 3D animated expressions to express their emotional states to other users.	6 sessions over 22 days	30-40 mins each session	Ppt views animated problem-based scenario. Scripted questions designed to elicit emotion-based responses. Ppt provides responses via the expressive avatar, text and speaking. Appropriate responses rewarded, inappropriate responses corrected. Researcher facilitates participant use.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Cheng, Huang, 2012	Examine whether the Joint Attention Skills Learning (JASL) System with data glove improves joint attention (JA) skills (pointing, showing, sharing and interaction).	Desktop computer, two projectors, a projection screen and data glove.	Observational learning, practice, reward reinforcement, error correction	Playroom environment		6 sessions over 6 days	30-40 mins	12 true/false and multiple-choice situational questions to learn JA concepts and practise skills in JA. Verbal question, 3D animated social event, prompting instructions, practice event, feedback.	None
Cheng, Huang, Yang, 2015	(i) to evaluate the effectiveness of the 3D Social Understanding system to enable ppts with ASD to learn non-verbal	HMD, virtual view controlled by keyboard and head-position sensor, joystick or cursor control	3D social modeling, promotion, and reinforcement rewards (Rebecca & Candice, 2010).	Bus stop, classroom environment	Full details not provided. (Examples given: putting your hand up to ask a question in class, what to do if a classmate speaks loudly to you in class, waiting in line when buying a	3 baseline sessions, 5 intervention, 3 maintenance sessions.	30-40 mins, once a week over 6 weeks	3D animated social models provided in different events. Problem-based social questions asked in True/False or Multiple choice format. Responses provided verbally or with keyboard. Immediate text feedback message provided. Each question permitted 3 times and hint provided if	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
	communication, social initiations, and social cognition; (ii) to assess the impact of using immersive equipment.				drink, what to say to someone who helps you)			incorrect response provided. Breaks given when emotional instability was observed.	
Cheng, Luo, Lin, Yang, 2018	Evaluate learning effects of the 3CFER system on recognition of complex emotion in people with ASD, and explore the personal phenomenon of using the system	Tablet - flat screen	Presenting animated emotions and social situations, ppts provide response, immediate verbal and written feedback provided.	Not stated	Stage 1: Identifying complex emotions of animated character with immediate feedback; Stage 2: Identifying a potential appropriate emotion from a social event situation; Stage 3: choosing a possible social situation causing a presented emotion in an animated character.	3 over 21 days	Self-paced (approx. 40 mins)	Social situation presented and animated emotions. Questions asked about scenario, ppts provide click responses, immediate feedback given	Not described fully. Paper-based emotion pictures and social stories

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Cheng, Ye, 2010	(i) examine feasibility of Collaborative Virtual Learning Environment (CVLE) for teaching social competence with children with ASD; (ii) examine whether the CVLE provides benefits of social competence (i.e. social understanding, social cognition and interaction) to children with ASD.	Laptop, Windows XP, 3D environment developed using 3D Max, Virtools and Poser. Avatars.	CVLE - Social Interaction System - Facilitated practice of social situations developed from a series of stories (Howlin, 1999). Reward/Correction and feedback on responses. Avatars to represent themselves.	A classroom, and an outdoor scene.	Ppts required to understand verbal and non-verbal communication, social understanding and expression. 3 out of 12 possible SSP questions/scenarios selected at random.	5 sessions over 18 days	30-40 mins	Ppt views animated problem-based scenario and listens to/reads questions. Scripted questions designed to elicit emotion-based responses. Ppt provides responses via the expressive avatar, text and speaking. Appropriate responses rewarded, inappropriate responses corrected. Researcher facilitates ppt use. SSP followed intervention session. Three SSP questions randomly selected from 12 in each situation.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Didehbani et al., 2016	(i) to assess the feasibility of a ten session virtual reality - social cognition training (VR-SCT) intervention for children with ASD and (ii) to measure changes in affect recognition, social attribution, and executive function	Second Life (Linden Lab, 2003) 3D virtual world, flat screen monitor, keyboard and mouse	Virtual Reality-Social Cognition Training (VR-SCT). Role-play social situation practice, feedback.	A school classroom, a school lunchroom, a playground, a campground, a race-track, a fast food restaurant, a technology store, an apartment, a coffee house, a sports store, and a park.	Examples: dealing with a bully, bonding with friends, confronting conflict, consoling a friend, handling social dilemmas (i.e. meeting a stranger, catching someone cheating)	10 sessions, over 5 weeks	1 hour, 2 x per week.	3 scenarios in each session. Social context provided to practice social communication and social cognition skills. Each social scenario has a targeted social learning objective. 'Coach' clinician describes and moderates each session and provides individualised feedback via avatar. Confederate clinician plays various parts in social interactions via avatar. Pre-established social prompts, and hints. Ppt chooses how to respond. 10 min social exchange, 5 minute feedback, redirection and suggestions. Clinician asks questions about ppt's awareness of social situation and discusses responses.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Herrera et al., 2008	(i) examine improvements in ppt's knowledge and autonomous management of real objects, (ii) examine improvement in pretend play and understanding of imagination	Flatscreen monitor (touchsreen)	Video modelling of functional and symbolic use of objects, object use practice in VR.	Not specified	Use objects in a functional, then symbolic - structured, prompted and symbolic - free play way.	28 sessions (approx 3 sessions per week)	20-30 mins	Progressive stepped teaching and exercises, from physical manipulation of objects (functional play) to imaginary play practice (symbolic). Multiple stages: Buying objects in a supermarket (e.g. trousers), 'functional use' video display of object use, video of child playing with miniature version of object (e.g. dressing a doll with toy trousers). Later stages: video display of character using representational objects in symbolic way (e.g. 'acting as if' a piece of cloth were a road). Imaginary transformation is shown explicitly in VR with a think bubble, transformation then shown without think bubble (as if occurring 'magically' in reality). Final stage: creative use of real objects (not detailed in paper).	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Ip et al., 2018	To examine whether VR intervention enhances emotional and social adaptation skills in children with ASD	Four-sided immersive Cave Automatic Virtual Environment (half-CAVE)	Emotion control and relaxation strategies; 4 simulated social situations; facilitation of consolidation and generalisation	Scenario 1 - home scene; Scenario 2 - Taking school bus and an education class at school; Scenario 3 - the school library; Scenario 4 - The Tuck Shop. Consolidation & generalisation - physical education class in the playground.	1 - Practising morning routine on preparing to go to school - Waiting patiently for bathroom, hand washing, brushing teeth; sequencing; 2 - greeting the teacher, handing in homework, following teacher's instructions, joining learning activities; 3 - Follow library rules, challenged by peer's inappropriate behaviours, queuing and checking out a book; 4- challenged by two emotion cognition situations - their chosen snack is sold out, a student jumps the queue.	2 sessions per week for 14 weeks (28 sessions)	40 mins (10 mins direct exposure, 30 mins observing)	Group therapy approach - Groups of 3-4 children at a time (10 mins direct exposure with VR, observing others rest of time); briefing, VR-enabled training, debriefing. Training consists of trainer guiding ppt through VRE of simulated scenes. Observers complete a workbook to encourage attention and concentration. Consolidation and generalization: social and coping skills learned are tested for generalisation - 2 scenes -1st child interacts with scenario alone with limited assistance, then trainer helps ppt to reflect on responses.	Waitlist control

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Kandalaft et al., 2013	Assess feasibility of VR - Social Cognition Training (VR-SCT) intervention in adults with HFA, and quantify social change over time.	Second Life (Linden Lab, 2003) 3D virtual world, flat screen monitor, keyboard and mouse	Virtual Reality - Social Cognition Training (VR-SCT)	An office building, a pool hall, a fast food restaurant, a technology store, an apartment, a coffee house, an outlet store, a school, a campground, and a central park	Examples: meeting new people and friends, initiating conversation with a roommate, roommate conflict, negotiating financial or social decisions, job interviewing, working with others and managing conflict, celebrating with a friend, consoling a friend, blind date, interacting with someone from a different background. Objectives: Emotion recognition and gauging interest, conversation skills and responding to others, theory of mind and self-assertion.	10 sessions, over 5 weeks	1 hour, 2 x per week.	Opportunities to engage in, to practise and receive feedback on meaningful young adult social scenarios. Different scenarios have different learning objectives. A coach in real life facilitates each session with the ppt. Confederate clinician appeared as VR avatar in each scene. Individualised feedback after each scenario. Ppt to incorporate feedback into following scenario.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Ke, Im, 2013	(i) determine the potential impact of VR intervention on social interaction and communciation in ASD children; (ii) determine which specific features and tasks of the VR intervention promote children's social interaction performance	Second Life OpenSimulator (OpenSim)-based VR - flat screen desktop computer	Role-play social situation practice. Prompting and reinforcing from facilitator.	A school environment, including playground and school cafeteria, and a child's birthday party.	Three social interaction tasks: (a) recognizing body gestures and facial expressions of a virtual communication partner, (b) responding and maintaining interactions at a school cafeteria, and (c) initiating and maintaining interactions at a birthday party.	6 to 9 (2 to 3 per task depending on child's rate of progress)	1 hour	One hour orientation session - interaction tasks explained, practice of navigation and communication in Second Life. One facilitator sits with child providing technical help and guidance on communication. Other facilitator interacts via VR avatar with ppt. Role-play scenario-contextualised interaction and situation-specific, naturalistic social communication. Facilitators follow semi-structured protocol, initiating, prompting and reinforcing interaction.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Ke, Lee, 2016	(i) Examine how the children with HFA proceed with the VR-based collaborative design task and interact with others in the virtual world; (ii) examine what specific processes of the VR potentially foster flexibility, identity- and norm-construction in children with ASD.	Second Life OpenSimulator (OpenSim)-based VR - flat screen desktop	Role-play social situation practice. Prompting and reinforcing from facilitator. Social interaction practice through collaborative architectural design and construction project.	Virtual neighbourhood	Collaborative problem-solving architectural design task. Ppt 'design team' need to agree on design plan for buildings in a virtual neighbourhood, the location of buildings, the sequence of construction and the style of architecture. Social interaction required in a meaningful context.	12 sessions scheduled (ppt 1 completed 11, ppt 2 completed 8, ppt 3 (TD) completed 12).	Approx 90 mins	During a design session, at least one ASD ppt and the TD ppt and two adults log into the virtual world. One adult plays role of expert architect who provided design comments. Other adult facilitator provides technical support, and prompts social interactions and communication and task focus. Structure of sessions not reported.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Lorenzo et al., 2016	to examine the effects of an immersive virtual reality system (IVRS) designed to develop and improve emotional skills in children with ASD	L-shaped screen ('semi-cave'), face recognition software to detect mood of participant and adapt the scene accordingly with robot arm eye-in-hand camera.	Social stories, role-playing	A party, a classroom	Going to a party, wanting to play with other children in a park, standing in line to go into a classroom, listening to a story told by a teacher in class, approaching children playing football, going on a fieldtrip with classmates, medical check-up at school, playing hide and seek with friends, sitting down next to some children talking, working in class and needing to ask for help. Ppt to identify situation and implicit emotions in different social stories and provide appropriate emotional responses.	4 sessions monthly for 10 months (40 sessions in total	35 mins each session	Social situation presented, facilitator explains task instructions and behaviour guidelines. Facilitator provides models, explanations and alternatives for a given situation. Ppt to provide answers to questions about the situation and implicit emotions in each social story and provide appropriate emotional responses. Facilitator provides feedback, and avatars respond	Same protocol but with 'classic desktop VR software application' instead of L-shaped VR (details not clearly specified).

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Lorenzo et al., 2013	(i) assess improvements in executive functions and social skills in ASD children in school environments with the IVE; (ii) to assess improvements relative to the learning difficulties of ppts; and (iii) evaluate the transfer of learned skills from the IVE to the real world	L-shaped screen (semi cave) 3D CristalEyes Stereography glasses, high quality audio and precise positioning system with Kinect sensor system, and four cameras. VE designed with Vizard.	TEVISA - Instructional and supportive tasks. Role-play.	(i) a school playground, (ii) a bedroom	Examples include: asking the teacher when ppt does not fully understand, planning and preparing work for the following day, underlining important information in text, managing an unscheduled change (e.g. absent teacher), managing unstructured (play) time with classmates, joining a line to enter a class, establishing a topic of conversation, inviting a friend to play at home.	80 sessions (2 per week)	25 mins per session	Variety of tasks to develop executive functioning and social skills. 16 TEVISA structured supportive tasks tasks completed five times each. Tasks involve resolving 'problems' typical of demands in school situations. They involve identifying and understanding the task/demand, and identifying steps to plan and carry out specific actions, and applying instructions in school environment. Instructions provided by avatars. Further details not provided.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Maskey, Lowry, Rodgers, McConachie, Parr, 2014	To examine whether using a CBT approach with an immersive VRE reduced specific phobia or fear for young people with ASD.	Immersive / wrap-around 'Blue Room' virtual reality environment	CBT with graduated exposure in VRE	Environment developed on an individual basis: Roadside scene, playground with pigeons, newspaper sales kiosk, bus stop, supermarket, car travelling through city, walking over bridge, classroom scene.	Tackling phobia / fear in situation specific to child	1 preparation session at home, 4 VR exposure sessions (2 sets of 2 sessions)	20-30 mins each session	CBT techniques introduced: identifying feelings, introducing 'feelings thermometer', relaxation techniques, deep breathing, positive coping thoughts. This training consolidated in Blue Room setting. In Blue Room intervention sessions, relaxation scene played, with breathing and stretching exercises. Graded exposure through VRE to situation to tackle (individualised to child). Scene repeated as many times as necessary, therapist facilitated use of anxiety management techniques learned to decrease anxiety at each level. Scene gradually increased in challenge over the four sessions. Parents observed sessions via video link. Family and therapist planned real life graded exposure to situation.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Mitchell, Parsons, Leonard, 2007	(i) Determine usability of VEs for teaching social understanding with ASD individuals, (ii) determine whether ASD participants show learning benefits of using the VEs	VE built with Superscape Virtual Reality Toolkit, run on laptop with Visualiser software. Use of joystick and mouse to move around and initiate interactions in first person perspective	Scaffolded learning and VR role-play. 4 levels of increasing social complexity. Encouragement and prompting of social convention understanding. Verbal and visual feedback.	Café scenario	Making judgements to choose an appropriate place to sit	2 sessions (on 2 separate days)	Approx. 40 mins (30-50 in range)	First, training activities to discover how to maneuver around VE. Then social judgements task: 4 levels of increasing social complexity (e.g. more people in scenario, more background noise). Ppts required to choose an appropriate seat and ask appropriate questions of VE characters. Visual and verbal feedback with suggestions and prompting with explanations to facilitate learning. Levels repeated if necessary.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Parsons, Leonard, Mitchell, 2006	(i) To gain insight from ppts with ASD whether there is merit to concerns raised about the use of VEs with autistic individuals, (ii) determine strengths of the approach from ppts' perspectives, (iii) determine whether ppts relate what they see to the real world	Laptop, joystick and mouse	Social skills training, role-play, facilitator providing feedback and explanation	Café and bus	Queuing and finding appropriate place to sit, and explaining reason for decisions.	3 VE intervention sessions	30-45 mins for cafe VE (session 1), 30 mins for bus VE (session 2), and 40 to 50 mins for both café and bus VE (sessions 3).	Training session (VE familiarisation). Then Cafe VE session: Scaffolded learning. 4 levels of increasing difficulty based on social complexity (e.g. no queue, quiet café, many empty tables to long queue noisy café, having to ask if they can sit with someone). Feedback provided and prompting. Level repetition encouraged if learning objectives not met. 2nd VE intervention session: Bus VE, similar to above. 3rd VE session: both cafe and bus scenarios presented, as above.	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Serret, Hun, Iakimova, Lozada, Anastassova, Santos, Vesperini, Askenazy, 2014	To verify the usability of JeStiMulE to improve emotion recognition in a hetergenous group of individuals with ASD	JeStiMulE - an interactive and multi-sensory virtual reality Serious Game (combines play and learning), on a gamepad, with avatar. Flatscreen desktop monitor	Computer game - trial and error feedback.	City environment, 5 different areas: a square, a garden, restaurant, theatre and a shop.	26 scenarios	Number of sessions varied by ppt according to performance. Played twice a week until JeStiMulE completed for four weeks maximum (8 sessions max)	1 hour per session	Nine expressions presented in game format (6 basic, 1 complex (pain) and 2 complementary non-emotional expressions (neutral and funny face)). Interactive and multi-sensory (visual, tactile and auditory stimulations) computer game. Designed for specific user requirements - adapted response options to all ppts. Each expression presented with a visual non-verbal code, a corresponding verbal written code and a tactile pattern. No verbal or written instructions - ppts discovered game intuitively or by trial and error. Leanring phase, then training phase. Designed for player and caregiver. Trained in recognising emotional expressions in static avatars and then dynamic ones. See each emotion ten times before	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
								progressing onto next game. Hierarchical learning - complexity increased with gradual additional emotinal cues and diversifying the avatars. Positive/Negative feedback provided with green/red flash. (In sum, a structured, prigressive and adapted learning procedure involving implicit learning, visual discrimination, attention to detail, categorisation and memory skills). Training phase: Ppt's avatar interacts with others in scene, at end of scene ppt required to identify emotin expressed by recognition or deduction. Reward provided in form of gaining piece of puzzle ppt is required to complete. Probabilistic progression.	

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Smith, Ginger, Wright, Wright, Taylor, Humm, Olsen, Bell Fleming, 2014	To test the feasibility and efficacy of interactive virtual reality role-play simulation, ‘‘virtual reality job interview training’’ (VR-JIT) to improve job interviewing skills in a sample of adults with ASD.	Desktop, speech recognition software	VR-JIT (VR role-play simulation). Involves repeated practice, hierarchical learning across progressive degrees of difficulty, and a reward system to reinforce behaviour change. Non-branching interviewing logic.	Job interview. Eight possible employment positions: cashier, inventory worker, food service worker, grounds worker, stock clerk, janitor, customer service representative, and security.	Job relevant content: Conveying oneself as dependable, as a team-player, professional and able to negotiate a schedule. Interview performance: sharing things in a positive way, sounding honest, sounding interested in the position, establishing overall rapport with interviewer.	5 visits (within 2 weeks)	10 hours over approx 20 trials	Didactic e-learning guidance on how to perform job interviews successfully, before simulated job interview experiences. Provides repeatable VR interviews; offers in-the-moment feedback with on-screen job coach, displays scores on key dimensions of performance, allows review of transcripts colour-coded for appropriateness of response. Hierarchical learning promoted with 3 difficulty levels determined by brusqueness of interviewer. Interviewer's demeanor and personality is variable (emotional realism).	Treatment as Usual

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Smith, Fleming, Wright, Losh, Humm, Olsen, Bell, 2015	(i) to evaluate the vocational outcomes of adults with ASD who previously completed the efficacy study of VR-JIT (Smith et al. 2014 above).	Desktop, flat screen monitor	As above	As above	As above	As above	As above	As above	Treatment as Usual (2 week waiting period between baseline and follow-up)

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Stichter, Laffey, Galyen, Herzog, 2014	Examine (i) the impact of iSocial on social competence of students with HFA, (ii) the degree of fidelity of the iSocial VLE to the SCI-A curriculum, and (iii) the degree to which students, teachers and parents find the experience socially valid	iSocial 3D Virtual Reality Environment delivered over the internet to PCs with high graphics card specifications.	iSocial 3D VLE Distance Education programme that implements Social Competence Intervention curriculum (SCI-A) - based on ABA, CBT and scaffolded learning (see Stichter, 2010 for full description)	Examples stated: collaboratively building a restaurant, and collaboratively helping to find a King's missing items and return them to him in a medieval world.	Not specified	31 lessons in 5 units. Each unit delivered over 2 weeks.	31 to 45 mins each lesson	University-based educator to guide ppts through Sci-A curriculum. Curriculum includes (1) facial expression recognition, (2) sharing ideas with others, (3) turn taking in conversations, (4) recognising feelings of self and others, and (5) problem solving. Involves (a) reviewing a previously learned skill (b) skill modelling, (c) skill practice in structured and naturalistic activities, (d) closing activity/review. Scaffold approach used	As above

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Strickland, Coles, Southern, 2013	To evaluate effectiveness of a web-based VR intervention to develop verbal (content) and non-verbal (delivery) interviewing skills	Desktop flatscreen monitor	JobTIPS multimedia training program (www.Do2Learn.com/JobTIPS). Involves step-by-step explicit instruction, video modeling, VR role-play, social narratives, video quizzes, repeated practice, feedback on performance, and systematic transfer of the skill to new conditions to promote generalization.	Office environment - interview office desk scenario	Provide responses to standard interview questions and situational questions	1	30 mins each session	Clinician assumed avatar role of interviewer. Provided feedback on first pre-intervention interview performance. the clinician presented the question, provided concrete explanations as to why one response might be more desirable than the original (from the perspective of the employer), and engaged the participant in repeated rehearsal opportunities across which guidance was gradually faded. Clinician also provided feedback and instructions for improvement on the ppt's non-verbals (e.g. body language, expressions, hand shaking).	None

Paper	Aims of study	Type of VR and other technology	Intervention approach	VR Environment(s)	VR Intervention Scenario(s) / Task Aim	No of sessions	Duration of sessions	Intervention components	Comparator intervention
Yang, Allen, Abdullahi, Pelphrey, Volkmar, Chapman, 2017	(i) investigate whether a pre-treatment biological motion fMRI task can predict therapeutic response to VR-SCT in adults with ASD; (ii) examine whether behavioural changes in emotion-recognition ability and theory of mind ability occur following VR-SCT.	VR-SCT training technology (Virtuak Gemini) over internet and on Windows computer. Facial emotion tracking using Faceshift Studio software. Webcam. Headphones with built-in microphone. MorphVox for voice-modulation	VR-SCT - a hierarchical strategy-based immersive role-play intervention programme to strengthen socio-emotional processing and socio-cognitive reasoning abilities. Interaction between coach clinician and participant is via avatar characters	Not specified	Dealing with confrontation, job interview, blind date	Two 1 hour sessions per week, over 5 weeks (10 hours total)	1 hour per session	First 3 sessions targeted learning recognising others, responding to others and self-assertion. Seven following sessions focussed on integrating strategies across varied and complex social situations. Ppt engages in a semi-structured live conversation with confederate clinician (CC). CC change character, conversational and emotional style to practice. Following practice conversation, coach and ppt engage in feedback discussion, including self-rating. Ppt then given further conversational opportunity to integrate discussed feedback	Control - no intervention

3.3. Outcomes and Results

Table 3 below provides brief summaries of outcome measures used in the included studies and results of primary outcome measures.

The following section summarises outcomes and results from the papers with respect to statistical significance of main findings or trends in the data, bearing in mind results from the quality assessment. Summaries draw together relationships between studies included in the systematic review with respect to intervention focus, with reference to intervention type and outcome measures.

Nine of the included studies did not perform statistical analyses on the data collected and either provided descriptive statistics only due to very small sample sizes or provided qualitative descriptions. Fifteen studies provided statistical analyses of their data.

3.3.1 Joint attention

Two studies examined the effect of intervention on participants' skills in detection and initiation of joint attention cues (Amaral et al., 2018; Cheng & Huang, 2012). Amaral et al. (2018) reported no significant differences in skills following the VR intervention; that is, no fixation on target object of joint attention preceded by fixation on the avatar's face (JAAT_Face) or not preceded by fixation on avatar's face (JAAT_NoFace). However, they do report a small, non-significant trend. Cheng and Huang (2012) on the other hand demonstrate improvements in all three of their participants, though no statistics are reported given the small sample size. The studies differed considerably however, and whilst the Amaral study explored eye fixations on target objects only, the Cheng study additionally explored effects on the child's initiation of joint attention cues (such as pointing and showing) with the use of a data glove. Nevertheless, the Amaral study performed relatively high on the Quality Assessment compared to the Cheng and Huang study, which scored 64% and 32% of the total score applicable to each study respectively on the Quality Assessment (see Appendix C for Quality Assessment scores for each study). However, it is possible that the low score for the Cheng study is due to the poor reporting of study details, rather than poor conducting of research.

3.3.2 Job interviewing

With respect to included VR studies for job interviewing skills (Burke et al., 2018; Smith et al., 2014; Strickland et al., 2013), all three papers report statistically significant results from the VR intervention. The efficacy results of the Smith study demonstrated that when compared to the

TAU group, the VR-JIT group showed improvement in interviewing scores across trials and increasing levels of difficulty; demonstrated significant improved skills in job interviewing with moderate to large effects; and had increased confidence in interviewing. The 6 month follow-up survey (Smith et al., 2015) showed a trend towards VR-JIT leading to positive employment gains compared to the TAU group. With the VITA practice system, Burke and colleagues (2018) demonstrated that participants enhanced their interviewing skills from the first VITA session to the final real face-to-face session by improving the manner with which they respond, identifying their individual strengths, engaging in self-promotion and self-advocacy and responding to social questions better. The Strickland study demonstrated improvements in the performance of participants in both the content and delivery of responses from first to final real-person interview session.

The Smith and Strickland studies were two of only five RCTs included in the systematic review. Though all three used customised evaluative measures involving ratings of interview performance, all three studies employed blinded raters. The three studies were rated amongst the top scorers on the Quality Assessment, with the Burke and Strickland studies having the highest scores of the three (79% and 78% respectively, and the Smith study attaining 69%). The quality of the studies therefore provides greater weight to the positive reported effects of the VR interventions used.

Although the studies focus on job interviewing skills, these skills draw on more general communication and social interaction skills, though they are specific to a particular context. Other studies examined effects on social skills in a more general context.

3.3.3 Affect recognition and social intentionality

In an RCT, Cheng et al. (2018) employed animated characters expressing emotions within contextual social situations. They reported significant differences between intervention and control groups at post-intervention following VR training in the identification and understanding of emotional expression in context, with the use of a customised Complex Emotion scale.

Lorenzo et al. (2016) employed VR-adapted Social Stories and role-play with feedback and recorded frequency of appropriate and inappropriate responses. The study demonstrates a statistically significant difference in identification of emotions and appropriate emotional responses in a real-world school setting, as rated by their teachers, over the period of ten months duration in which the intervention was provided. Whilst two interventions delivery types are provided – an immersive L-shaped semi cave (IVRS) and desktop VR, both showed

significant improvements in socio-emotional responses, but the IVRS yielded greater improvement over the desktop VR system.

Serret et al. (2014) employed a multi-sensory hierarchical 'serious game' learning format (JeStiMulE) for participants to develop emotion recognition skills implicitly, without verbal or written instructions or feedback. Outcome data was collected from performance on the VR computer game, which yielded statistically significant improvement in emotion recognition following the JeStiMulE program. This significant improvement was also shown to generalise from animated avatars to real-life characters in the program.

Whilst several studies employed different intervention programmes, three studies (Didehbani et al., 2016; Kandalaft et al., 2013; Yang et al., 2017) all employed a VR-SCT immersive role-play, hierarchical learning and feedback intervention, and measured performance pre- and post-intervention. Didehbani reported significant improvement in emotion recognition from pre-to post-intervention as measured by the NEPSY-II Affect Recognition test following training. Both Yang and Kandalaft employed the ASC-Social Perception test to measure affect recognition and non-literal language interpretation. They found significantly improved performance post-VR-SCT training with moderate effect sizes on the ASC-SP. All three studies additionally employed the Triangles test to measure theory of mind and social intentionality and each demonstrated significant improvement with small to moderate effect sizes on test post-intervention.

The Cheng, Serret, Didehbani, Kandalaft and Yang studies gained relatively high scores on the Quality Assessment tool (percentage score of total applicable points: 69%, 71%, 79%, 71% and 78% respectively), though the Lorenzo (2016) study scored relatively poorly in comparison (50% of total possible applicable) in part due to issues with clarity of reporting.

3.3.4 Socio-cognitive reasoning and social decision making

Two papers examined socio-cognitive reasoning and decision making (Mitchell et al., 2007; Parsons et al., 2006). Mitchell and colleagues used scaffolded learning and role-play in VR with the aim for participants to develop understanding of social conventions when choosing a seat in a café and on a bus. This was assessed through video recordings of sessions by blind raters. To rule out effects due to practice effects rather than VR experience, half the participants experienced the VE between video measures at Times One and Two, whilst the other half experienced the VE between video measures at Times Two and Three. Results demonstrate significant gains in both seat choice and social reasoning directly following VE than not. Parsons and colleagues (2006) conducted a qualitative study of two participants' experiences in using the same VR scenarios as in the Mitchell study. The ASD participants demonstrated reduced

social errors in later VR sessions compared to earlier ones, and knowledge was transferred across settings (café to bus scenario). However, there was variability between the participants and one struggled with asking questions of strangers. Both participants commented that the experience had helped them in real life getting on buses and the tube. The Mitchell study scored moderately well on the Quality Assessment (65%) whilst the Parsons study scored poorly (31%), thereby providing some but limited support for VR intervention in this area.

4.5 Executive functioning

Didehbani et al. (2016) and Lorenzo et al. (2013) both examined VR intervention effects on executive functioning. Didehbani employed Second Life software with a desktop computer and used VR-SCT (role play social situation practice), and used the NEPSY-II Auditory Attention, Response Set and Analogical Reasoning subtest of executive functioning. Whilst their study yielded positive significant results in relation to emotion recognition and social attribution (described above), the measures of executive functioning were all non-significant.

Lorenzo employed an L-shaped Semi Cave VR environment for role-play social scenarios typical of school social demands. Behaviours were measured through systematic and structured observational analysis. Although no statistical tests were carried out, descriptive statistics of frequency averages of target behaviours yielded a trend in improvement in relation to executive functioning. Unfortunately, both studies by Lorenzo and colleagues (Lorenzo et al., 2016, 2013) employing the L-Shaped semi cave IVE do not clearly explicitly differentiate between changes of target behaviours that relate to executive functioning and social skills or emotion recognition and appropriate social responsiveness, but rather group them together. As such it is difficult to interpret their data accurately. The two papers score relatively poorly on the Quality Assessment (Lorenzo et al., 2013 - 48%, Lorenzo et al., 2016 – 50%).

Stichter et al. (2014) delivered an iSocial distance learning programme for developing social competency skills as well as executive functioning skills with the use of scaffolded learning, as well as CBT and ABA techniques. Parental reports of their children's executive functioning at home, using the BRIEF assessment, demonstrated significant improvement from pre-to post intervention. However, teachers' reports of the same at school did not show significant change.

4.6 Emotional regulation

Ip et al. (2018) and Maskey et al. (2014) employed similar VR environments and had overlapping intervention content, though with somewhat different aims. Both Ip and Maskey employed an

immersive VR cave environment to deliver their intervention; Ip to enhance emotional and social adaptation skills, and Maskey to tackle individuals' fears in specific situations. Both incorporated relaxation strategies into the intervention for emotion control purposes; Ip employing this technique whilst providing opportunities to practise difficult-to-manage social situations at school; and Maskey employing it within a more formal CBT graded exposure approach to fearful situations.

Ip employed the PEP-3 Affective Expression subtest to assess emotion expression and regulation with significant improvements from pre-to post-training, as well as the PEP-3 Social Interaction subtest to examine social behaviours and skills, again with significant positive changes. Maskey also demonstrated positive improvement for the children in their study in terms of anxiety (with the SCAS) and in terms of behaviour change in tackling the situation (measured by comparison of behaviour vignettes). They demonstrated clinical reliable change for participants, with all but one child being responders to treatment and half the participants overcoming their phobia completely in real life. The Ip study scored moderately well (63%) on the Quality Assessment, and Maskey scored amongst the highest of included studies (79%), lending greater weight to the support for the utility of such interventions for ASD participants for emotional regulation purposes.

4.7 Social competency skills – communication and interaction

With respect to more general social communication, conversation and interaction abilities, several studies employed VR with the aim to develop such skills. Many of these studies employed very small sample sizes and as such do not provide statistical analyses but descriptive statistics of the results.

Cheng, Chiang, Ye, & Cheng (2010); Cheng, Huang, & Yang (2015); Cheng & Ye (2010) all employed similar study designs with just three participants, using multiple probes and multiple baselines. The Cheng et al (2010) study examined understanding of and expression of empathy, and employed a customised Empathy Rating Scale with three sub-tests examining kindness, tolerance and respect. Cheng et al. (2015) examined socially appropriate behaviours of social initiation and social cognition using a customised Social Behaviours Scale. Cheng and Ye (2010) employed a customised Behaviours Checklist to examine social understanding, social cognition and interaction. Each of the three studies employed VR as a means to provide facilitated practice of various social situations with response reward or correction and feedback. All three studies demonstrated a trend in their three participants towards improvement in each of their respective target areas of intervention. The improvement was shown to continue not just from

baseline to intervention but through the maintenance period also. The studies however, were rated relatively poorly in comparison to other studies on the Quality Assessment (Cheng et al (2010) – 44%, Cheng et al. (2015) – 36%, and Cheng and Ye (2010) – 40%).

Ke and Im (2013) examined Second Life VR intervention effects on social interaction and communication skills in children with ASD with the use of facilitated role-play, prompting and reinforcing. Ke and Lee (2016) employed a similar approach with the same VR type but through a collaborative design task with other peers with the aim to foster interaction flexibility. Both studies again involved small sample sizes and the studies report individuals' results. Ke and Im (2013) report an increase in successful performance for their four ASD participants in many areas, including initiating, responding and maintaining interactions, greetings and ending conversations, although turn-taking skills were variable across participants. Ke and Lee (2016) provided results from two ASD participants, and demonstrated an improvement in social interaction flexibility in terms of seeing the other person's perspective and communication. However, participants' performance in terms of identity and norm construction expressed through collaborative design fluctuated and varied across the two participants. The Ke and Im, and Ke and Lee studies attained scores of 58% and 42% on the Quality Assessment, thus performing relatively low to moderate in terms of risk of bias.

Studies which performed much better on the Quality Assessment and examined social interaction and communication skills included Stichter et al. (2014) (75%) and Ip et al. (2018) (63%). Stichter et al. (2014) delivered an iSocial distance learning programme for developing social competency skills, using scaffolded learning, CBT and ABA techniques. Parents reports of their children's social competencies at home (with the SRS) demonstrated significant improvement from pre-to post intervention. However, teachers' reports of the same at school did not show significant change. As well as intervention effects on emotional expression and regulation, as described above, Ip et al (2018) additionally measured social interaction behaviours and skills with the PEP-3 Social Interaction subtest. They report statistically significant changes from pre- to post-intervention and positive significant differences from the control group.

The studies included in the review vary widely in terms of target skills and behaviours for intervention (from joint attention and pointing to interviewing skills). Some of these target skills could be considered more low-level or fundamental building blocks of social interaction, on which higher-level interaction and communication skills build. For instance, joint attention is considered to be one of the first building blocks of social interaction. Joint attention involves

quite a complex coordination of eye contact, gaze monitoring, showing, pointing and point-following between two individuals and the sharing of attention toward an object or event (Bakeman & Adams, 1984). This ability develops between 6 and 12 months of age in typically developing children, and it is argued to be critical to the development of other social and language abilities (Charman, 2003).

It is worth noting that the systematic review indicates poor support for the use of VR intervention in the development of joint attention skills; whereas more higher-level cognitive skills, such as conversational interchange, skills can improve through VR training. It is unclear from the research conducted whether the individuals who experienced improvements in higher-level social abilities, such as conversational interchange skills, already had good joint attention abilities as they were not assessed prior to intervention. However, it is possible that the results from the review imply that a critical period occurs during which such fundamental joint attention skills can develop, and are unattainable or much harder to develop thereafter despite any intervention/training. In contrast, higher-level social skills can develop with training. Further empirical research is required to determine whether some skills are necessary and essential in order for others to develop. Furthermore, it is yet to be determined why the development of joint attention skills is impaired in ASD or how joint attention is related to later social skill development, and are thus important future areas of research focus.

Table 3. Outcome measures and results

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Amaral et al., 2018	Primary: Customised Joint Attention Assessment Task (JAAT). Secondary: Autism Treatment Evaluation Checklist (ATEC); Vineland Adaptive Behaviour Scales (VABS); Profile of Mood States (POMS), Hospital Anxiety and Depression Scale (HDS) and Beck Depression Inventory (BDI)	JAAT - to assess detection of initiation of joint attention cues Unit of Measurement: number of items ppt accurately identified from avatar's action cues; ATEC - to measure treatment effectiveness for autism symptoms and improvements; VABS - to assess adaptive functioning, POMS, HADS and BDI to assess mood, anxiety and depression	**Between baseline and primary follow-up:** JAAT_NoFace = Mean difference = 2.60 (CI = -2.20, 7.40). JAAT_Face mean difference = 2.87 (CI = -0.07, 5.80) (non-sig). **Between baseline and secondary follow-up:** JAAT_NoFace = Mean difference = 1.33 (CI = -4.47, 7.14). JAAT_Face mean difference = 3.13 (CI = -2.00, 8.27) (non-sig).	Not reported	Primary measure (JAAT) was customised and not clinically validated. Secondary measures validated.
Burke et al., 2018	The Marino Interview Assessment Scale (MIAS)	MIAS - measures degree to which ppt used skills in interview. Measurement: 1 to 5 scale. (1 = did not use strategy, 5 = accomplished strategy).	Mean MIAS score over all questions increased by 0.58 units (SE = 0.12, t (118) = 5.04, $p < 0.0001$). Mean scores on MIAS over all questions between first and final interview for ppts with autism increased by 0.49 units (SE = 0.15, t (84) = 3.35, p = 0.006).	Blinded researchers/assessors	No test-retest scores. No data regarding potential effects due to retesting, practice effects. Inter-rater reliability not tested.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Cheng et al., 2010	Empathy Rating Scale (ERS) (Lin, 2008).	ERS - measures Degree of Kindness (KE), Degree of tolerance empathy (TE) and Degree of Respect Empathy (RE). 5 questions chosen randomly from the ERS in each session. 4-point Likert Scale (0 = agree to 3 = strongly disagree).	Results derived from visual inspection of ERS scores from baseline through intervention and maintenance. Scores of the three ppts varied at baseline. Scores for Ppts A and B increased over intervention sessions. Ppt C's scores increased comparatively slower but relatively more stable. Ppt A's scores declined somewhat during the maintenance period, Ppt B's scores stabilised (KE and TE) and increased further (RE). Ppt C's scores increased (KE) stabilised (TE) and increased further during maintenance (RE) and had a higher score than Ppts A and B in maintenance period.	A researcher and a teacher	Discussed with professionals and autism experts for 'validity and suitability'

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Cheng & Huang, 2012	Joint Attention Skills Scale (JASS) with use of ten picture cards of joint attention.	JASS assesses initiating and responding joint attention skills (0-5 scale, total score = 40);	All 3 ppts showed improvement in areas of pointing, showing, sharing and interaction with the data glove from baseline to intervention and maintenance. No statistics reported.	One researcher and two observers (1 teacher, 1 researcher) observing video recorded sessions. Questionnaire administered to parents.	Discussed with professionals and autism experts for 'validity and suitability'. Inter-observer agreement > 97.5%.
Cheng et al., 2015	The Social Behaviours Scale (SBS). Ppt selected three random social events cards (observation of social events) and responded to 12 questions on the SBS.	SBS evaluated ppts' abilities to perform socially appropriate target behaviours: Non-verbal Communication, Social Initiation and Social Cognition. Measurement: 6 true (3)/false (0) questions, 6 MCQs (0 to 5 scale), SBS score based on appropriateness of response, maximum score = 48.	All three ppts demonstrated improvement in mean SBS scores during the intervention stage from baseline, and further improvement during maintenance phase. Ppt 1 Mean scores for baseline, intervention and maintenance respectively: 10, 22, 27.3; Ppt 2 Mean scores: 11.8, 23.5, 26; Ppt 3 mean scores: 8.7, 19.8, 25.	The researcher, an observer, and a teacher.	Discussed with professionals and autism experts for 'validity and suitability'. Average Inter-observer reliability of 97.4%, 97%, and 92.7% for the 3 ppts.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Cheng et al., 2018	Complex-emotion (CE) scale.	Assessed identification and understanding of facial expression in context. Four complex emotion-picture and situation-picture cards used.0-6 per question (max total 72).	Difference between groups at pre-test - ns. Sig difference between groups in the post-test session ($F(1,24) = 45.127$, $p= 0.000$. Pre-test scores: Experimental group M = 36.75, Control M = 36.50. Post-test scores: Experimental group M = 56, control group M = 41.5.	Researcher and two observers	Validity confirmed through assessment with ASD experts.
Cheng & Ye, 2010	Behaviours Checklist (BC), using Social Situation Pictures (SSP). Determined through observation of performance of ppts.	Evaluates social competences: (i) understanding of social interaction; (ii) social cognition and interaction in performing social behaviour (through use of expressive avatar). Measurement was percentage of appropriate answers.	Ppt A's appropriate social behaviours score increased from an average of 45.3% at baseline to 82% in the last intervention session, which was maintained during the maintenance period. Ppt B averaged 69.3 at baseline and increased to 90% at intervention, and remained at 85% at maintenance. Ppt C's scores increased from a mean of 56% at baseline to 86% in the last intervention session and averaged at 81.5% at maintenance. Similar results were obtained for Understanding of Social Interaction.	One researcher and one teacher (not blinded).	Measures discussed with professionals and ASD experts for their validity and suitability

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Didehbani et al., 2016	Developmental Neuropsychological Assessment (NEPSY-II) Social Perception - Facial Affect Recognition subtest (Korkman, Kirk, & Kemp, 2007); Triangles (Abell et al., 2000); NEPSY-II Auditory Attention: NEPSY-II Response Set; NEPSY-II Analogical Reasoning.	NEPSY-II Affect Recognition - test of facial affect recognition; Triangles - test of understanding of social intentionality; NEPSY-II Auditory Attention - test of selective attention and ability to maintain attention; NEPSY-II Response Set - test of complex ability to shift and inhibit previously learned responses.	NEPSY-II Affect Recognition: Paired t-tests comparing pre and post differences: significant increases t (24)= -3.40, p = 0.001. Social Attribution - Triangles Intentionality: t(23) = -2.28, p = 0.016 following training. Total Triangles score: t (23) = -1.93, p = 0.033. Attention and Executive Functioning: significant increases on analogical reasoning t (17) = -2.33, p = 0.016 following training. There was no significant change on NEPSY-II Auditory Attention and Response Set. There were no significant differences on secondary analyses examining differences between ASD only ppts and ASD and ADHD combined ppts.	Not stated.	NEPSY-II AR: high reliability coefficients (rs = 0.85 to 0.87) and moderate test-retest coefficients (rs = 0.50 to 0.58); NEPSY-II Auditory Attention and Response Set: both have high reliability coefficients (r = 0.81 to 0.88) and moderate to high test-retest coefficients (r = .53 to .84). NEPSY-II Analogical Reasoning: not reported. Triangles: high test-retest reliability r = 0.76 to 0.88 and concurrent validity r = 0.78 to 0.93 (Hu, Chan, & McAlonan, 2010). Inter-rater reliability for total score Kappa = 0.80 (p < 0.001), and for intentionality score Kappa = 0.83 (p < 0.001).

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Herrera et al., 2008	Structured and Non-structured Test of Pretend Play (ToPP) (Lewis & Boucher, 1997)	Observational assessment of functional and symbolic structured-play and free-play. Higher score if ppt responds to elicitation without modelling or instruction. Observations of quantity and category made. 1 to 34 points.	Both ppts had good functional use of objects pre-intervention. Ppt 1 showed a 6.5 point improvement in symbolic play from 29. 8 months of development to 42.8 post-intervention (structured), and 5 points improvement from 37.3 to 48.1 months (free play). Ppt 2 showed 4.75 points improvement, from 40.3 to 49.8 months level of typical play development (structured), and 5 points improvement, from 18 to 29.8 months (free play).	4 observers (2 independent of research team)	Inter-rater reliability: r = 0.877 (p = 0.004) between observers from the team and 0.923 (p = 0.001) between external observers; 0.838 (p = 0.009). between external and internal observers. Validity not reported.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Ip et al., 2018	**Primary:** Psychoeducation Profile Third Edition (PEP-3, Schopler, Lansing, Reichler & Marcus, 2004; Chinese version Skek & Yu, 2014); subtests of Affective Expressions and Social Reciprocity. **Secondary**: Faces Test (Baron-Cohen, Wheelwright, & Joliffe, 1997); Eyes Test (Baron-Cohen, Wheelright, Hill, Raste, & Plumb, 2001); Adaptive Behaviour Assessment System 2nd Edition (ABAS-II) (Harrison & Oakland, 2003) and Childhood Autism Spectrum Tests (CAST) (Williams et al., 2005) both completed by parents. Raven's Progressive Matrices (Raven and Court, 1998);	**Primary Measures:** PEP-3 Affective Expressions - emotion expression and regulation; PEP-3 Social Interaction - assesses social skills and behaviours in ASD. **Secondary:** Faces (Chinese model adaptation) - emotion and mental state recognition; Eyes - emotion recognition; ABAS-II - parent completed norm-referenced assessment of adaptive skills and functioning; CAST -parental questionnaire to screen for ASD. Raven's -measure of non-verbal ability.	Training group scored higher on emotion expression and regulation after the training (M = 20.2, SD = 3.00) than before the training (M = 18.9, SD = 3.57, t(35) = −2.174, p = .037). Training group also scored higher on social interaction and adaptation after the training (M = 21.8, SD = 2.99) than before training (M = 20.2, SD = 3.43, t(35) = −3.987, p < .0005). Control group: all measures non-significant, except parent reported ABAS scores of communication p = .03) and community (p = .04). Statistically significant interaction between group and timepoint on affective expressions, F(1, 70) ¼ 5.223, p = 0.025, partial η^2 = 0.069; significant interaction between group and timepoint on social reciprocity, F(1, 70) = 7.769, p = 0.007, partial η^2 = .100.	Not stated	Chinese version PEP-3 validated (Shek & Yu, 2014)

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Kandalaft et al., 2013	Advanced Clinical Solutions Social Perception Subtest (ACS-SP; Pearson 2009); Reading The Mind in the Eyes, (Baron-Cohen et al., 2001), Triangles (Abell et al., 2000); Social Skills Performance Assessment Version 3.2.(2) (SSPA; Patterson et al., 2001); Facial Expressions of Emotion Stimuli and Tests (Ekman 60; Young et al. 2002).	ASC-SP - measure of verbal and non-verbal emotion recognition; 'Eyes' - Theory of Mind; 'Triangles' - theory of mind and social intentionality; SSPA - Conversation Skills; 'Ekman 60' - measure of facial affect recognition of the basic emotions (happy, sad, fear, surprise, anger, disgust).	**Social Perception Total score**: significant improvement from pre-to post-testing ($t(7) = 2.83$, $p < .025$, $\eta^2 = 0.53$, mean difference (SD) = 2.00 (2.00), 95 % CI: 0.33 to 3.67. **SP-Prosody** score significantly improved after treatment, $t(7) = 3.27$, $p = .014$, $\eta^2 = 0.60$, mean difference (SD) = 1.63 (1.41), 95 % CI: 0.45 to 2.80. Ekman: significant improvement from pre-to post-treatment: $t(7) = 3.99$, $p = .005$, $\eta^2 = 0.69$, mean difference (SD) = 2.38 (1.69), 95 % CI: 0.97 to 3.78. Triangles Intentionality also improved, $t(7) = 3.45$, $p = .011$, $\eta^2 = 0.63$, mean difference (SD) = 1.25 (1.04), 95 % CI: 0.39 to 2.16. Other measures were non- significant.	SSPA - raters blinded to pre- and post- time points. Other tests - not stated.	ASC-SP - internal consistency - r = 0.69-.81, test-retest reliability coefficient - r = .60-.70, inter-scorer agreement - .98-.99. Normative scaled scores available for the 4 subtests. Eyes??; Triangles??; SSPA - Telephone survey - no validation.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Ke & Im, 2013	Observation of child's behaviour and communication in virtual world and reality. Frequency data on each target social interaction collected (rate per hour/session); Social Skills Questionnaire: Parent and Pupil Forms (SSQ; Spence, 1995); Assessment of Perception of Emotion from Facial Expression and Posture Cues (Spence, 1995).	Observation aimed to examine three social communication and interaction skills: (i) Responding and maintaining interaction;(ii) Leading or initiating interaction; and (iii) Non-verbal communication. SSQ Parent and Child versions - to assess social behaviour; Perception of Emotion - to examine children's ability to identity emotion from facial expressions and body postures in photographs.	No statistical analysis. Graphical / visual analysis. Individual data reported. Overall participants demonstrated an increase in successful performance trials during intervention sessions on Responding and Maintaining Interactions, Appreciation expression, Initiating Interactions, Greetings, and Ending Conversations. Turn-taking skills were variable across participants. All ppts demonstrated increase in Spence's facial expression and posture cue assessment and on the SSQ scores (both parent and child forms) from pre to post intervention.	2 coders (further details not specified).	All reported to have good internal consistency and used in prior studies of social interaction in ASD.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Ke & Lee, 2016	Qualitative time-series and micro-behaviour analysis. Frequency of behaviours recorded. Assessed Flexibility, Identity Construction and Norm Construction.	Flexibility: examines flexibility in social interaction (e.g. attention to other's perspective), communication (e.g. complexity of language), and design (e.g. comfort with ambiguity). Identity: examines self-identity representation via design, expression through group/community identity and explaining motives/conduct. Norm construction: examines following preset rules of the design task, and exploring and creating potential norms with others.	Flexibility: Both ASD ppts showed increasing trend of overall flexibility across sessions. Identity Construction: Both ASD participants showed fluctuating performance across sessions. Norm Construction fluctuated for ppt 1, but Ppt 2 showed an increase.	Not specified.	N/A

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Lorenzo et al., 2016	Frequency of adequate and inadequate responses, facial expressions and behaviours in each of the Social Stories. Observations divided into three sub-categories: a) behavior during the presentation and description of the social stories; b) emotional response when carrying out the social stories; c) compliance with the behavior guideline.	Data collected to infer changes in children's social emotional recognition and emotional responsiveness skills. Identification of situation and emotions: 1 to 6 points; Appropriateness of emotional response: Category from 0 to 6 (higher category = better performance).	Identifying the social situation: the mean scores increased in the IVRS group by 0.9 points .($p < .001$)compared to 0.1 points in the control group ($p = 0.37$) from first to last session. Final average grades for the IVRS and control group are statistically significant ($p = .03$), with better performance in IVRS group. Behaviour during presentation: Frequency of appropriate behaviour increased by 2.4 points in IVRS group ($p < 0.001$) compared to control group (increase of .2 points $p = .47$). Emotional response: increase of 4.4 points in IVRS group and 2.3 points in control group, both significant ($p < .05$) and statistically significant from one another ($p = .002$). Compliance with behaviour guidelines: IVRS group increase of 9.4 points ($p < .05$) compared to 6.4 points ($p < .001$) in control. Difference between both groups $p = .01$.	The facilitator	Not stated

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Lorenzo et al., 2013	Systematic and structured observational analysis of ppts' responses and behaviour in IVE.	Data collected to infer changes in executive functioning and social skills	**Primary school children**: Positive progress observed in score category from first (>75% in score category 1) to last session (average rates of between 41.1% and 54.6% in score category 4). **Secondary school children**: Similar results as in primary school children. Average % rates for categories not reported. Graphical reporting only. **Transferring IVE knowledge to real world classroom**: increase of 1.6 and 1.1. units for primary and secondary pupils respectively.	Researcher and school teachers for real world classroom.	Not stated.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Maskey et al., 2014	(1) Spence Children's Anxiety Scale-parent version (SCAS-P) and child version (SCAS-C). (2) Target behaviour vignettes created based on responses to standard format questions using protocol developed by The Research Units. (3) Children and parents rated their confidence in tackling the target situation at the beginning and end of each session, and in real life occasion.	SCAS-P and SCAS-C - to assess anxiety symptoms. Target behaviour vignettes - to record behaviour change over time for anxiety related to specific situation. Confidence ratings determined confidence / comfortableness at tackling situation. SCAS: 0-3 scale for 44 items. Target behaviour vignettes: 9 point scale (1 = normalised to 9 = disastrously worse). Confidence ratings: on scale from 0 (not at all comfortable) to 6 (very comfortable).	Pre-treatment: SCAS-P group Mean = 40.7 (SD = 14.3), SCAS-C 43.3 (15.0). 6 weeks post-treatment SCAS-P = 34.2 (17.9), SCAS-C = 35.5 (21.8). 6 months post-treatment: SCAS-P = 31.1. (15.9), SCAS-C = 29.6 (19.7). 12-16 months post-treatment: SCAS-P = 29.7 (19.8), SCAS-C = 27.9 (20.1). Reliable Change Index pre-post treatment at 12-16 months: SCAS-P = 1.6, SCAS-C = 1.94. Target behaviours: All but one child were responders to treatment (rated 3 or less). 4 children completely overcame phobia and effects were maintained at 12-16 months.	Vignettes compared by expert panel	**SCAS** has good criterion validity in typically developing children, and widely used in ASD studies, high internal consistency and good validity. **Target behaviour vignettes**: not stated. **Confidence rating**s: not stated.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Mitchell et al., 2007	3 video measure sessions. Five video excerpts (2 café scene, 3 interior bus scenes) to assess specific learning objectives of VE experience. Bus scenes used to assess generalisation of learning.	Ppts' social understanding inferred from their descriptions and explanations of behaviour in a scene. Ppts assessed on social appropriateness of action, and how social their reasoning/explanation for action was. 6-point bipolar scale (intermediate ranks undefined).	Significant gains in seat choice were more common in video sessions directly following VE than not directly following VE (McNemar χ^2 (1, n = 9) = 7.11, p < .01). Significant gains in social reasoning were more common in video sessions directly following VE than not directly following VE (McNemar χ^2 (1, n = 19) = 7.58, p < .01). Majority of significant VE-associated gains were apparent in subset of the ppts.	10 blind independent raters	Not stated
Parsons et al., 2006	Descriptive qualitative analysis of performance. Description of how well ppts tackled demands and their responses to scenarios, and whether responses changed over time with VE practice.	Examines social appropriateness of action and understanding of social norms.	Ppt 1: Social errors made in early VE sessions, but then absence of errors in later sessions (learning from mistakes was incorporated into VE practice).Incorporation of knowledge into next levels without prompt. Transfer of knowledge between contexts and between media. Generalisation of skills to real world at 3 month follow-up. Ppt 2: Similar responses as Ppt 1, but struggled with some social concepts and timidity in asking stranger questions continued in real-world.	Researcher	Not stated

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
			Ppts reported the VEs had helped them in similar situations in real world and reported why they had helped.		
Serret et al., 2014	Performance data collected from computer game	Quantified performance score is an approximate measure of emotion recognition skills	Ppts more accurate at recognising avatar emotions after JeStiMulE training (M = 64.65, SD = 17.50 versus M= 31.16, D = 13.34, $p < .001$). Significant main effect of session ($F(1,32) = 98.48$, $p < .001$). Ppts also significantly better at recognising emotions in real life characters after JeStiMulE (M = 61.21, SD = 28.83 versus M = 37.39, SD = 20.96, $p < .001$). Significant main effect of session ($F(1,32) = 49.09$, $p < .001$).	N/A	Not stated.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Smith et al., 2014	Primary measures: 2 standardised live role-play job interviews with 'standardised role-play actors' (SRAs), and the self confidence measure. Secondary measures: Psychosocial Interview, Social Responsiveness Scale 2nd edition (SRS-2), Vocational interview, neurocognitive assessment (RBANS) and social cognitive assessment (BLERT) and an emotional perspective-taking task, training experience questionnaire (TEQ) to evaluate feasibility of the VR-JIT.	Live role play interviews assessed (1) comfort level; (2) negotiation skills; (3) dependability; (4) sounding easy to work with (team-worker); (5) sharing things in a positive way; (6) sounding honest; (7) sounding interested in the position; (8) conveying professionalism; (9) establishing rapport with interviewer. Role play videos: Total score (range 1 to 5) across nine domains; Self confidence measure: seven-point Likert scale.	Significant group by time interaction found for the total role-play assessment scores ($F(1,21) = 4.4.$, $p = 0.046$, large effect in the VR-JIT group, no change in the TAU group ($d = 0.83$ and 0.11 respectively). VR-JIT group showed trend level improvement in interview content ($F(1,24)$ n=4.0, $p = 0.056$, $d = 0.76$), and interview performance ($F(1,24) = 3.2$, $p = 0.086$, $d = 0.73$). TAU group showed no improvement ($d = 0.08$ and 0.10 respectively). Self-confidence measure: A trend-level group-by-time interaction ($F(1,22) = 3.9$, $p = 0.06$), VR-JIT group effect size $d = 1.15$, TAU group effect size $d = 0.32$.	Role-play interview videos assessed by two raters (expertise as HR interviewers), blinded to condition	Reliability established by double scoring 20% of videos (ICC = .94), rater drift prevented by raters meeting after every 20 videos and reaching consensus.

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Smith et al., 2015	Instructed to complete a brief follow-up survey after 6 months. Seven questions to reflect on past 6 months, focussing on whether they had obtained paid employment or a competitive volunteer position.	To determine whether ppts had gained competitive employment (paid or volunteering) following VR-JIT.	No between group differences found at follow-up for most questions, except proportion of ppts who accepted a position was larger in VR-JIT group (53.3%) compared to controls (25%), but difference was only a trend ($p = 0.09$). Post-test self-confidence and prior paid employment were not related to accepting a position. Only observed improvement between pre-and post-test job interview role-plays was associated with completing more interviews for a position ($r = 0.55$, $p = 0.02$). Advanced social cognition was associated with accepting a competitive position at a trend level ($r = 0.37$, $p = 0.09$).	Researcher	Not stated

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Stichter et al., 2014	The Social Responsiveness Scale (SRS), Behaviour Rating Inventory of Executive Function (BRIEF). Both SRS and BRIEF completed by one parent and one teacher. Performance measures: Reading the Mind in the Eyes; The Faux Pas Stories, Diagnostic Analysis of Non-verbal Accuracy-2 Child Facial Expressions (DANVA-2-CF), DKEFS, Conner's Continuous Performance Test-II (CPT-II)	SRS evaluated social abilities in home and at school in 5 domains: social awareness, social cognition, social communication, social motivation, autistic mannerisms. 4 point Likert scale raw scores, BRIEF - used to rate executive function, - 8 subscales combined to create indexes of behaviour regulation (BRI) and meta-cognition (MI). 3 point Likert scale.	Parents' SRS report: significant improvement from pre- to post-intervention ($t = 3.72$, $p < .01$); Parents' BRIEF report at home: significant improvement ($t = 2.43$, $p < .05$). Teachers reports of SRS and BRIEF were non-significant. Task Performance measures were non-significant from pre-to post intervention, except on D-KEFS Design Fluency total measure ($t = 3.82$, $p < .01$)	SRS and BRIEF completed by one parent and one teacher. Researcher measured other outcomes assessed.	SRS subscales demonstrated as having acceptable internal consistency ($\alpha = .77$-$.92$), high construct validity ($rs = .65$-$.70$) and contrasted groups validity (can distinguish between TD and ASD children). BRIEF subscales demonstrated good internal consistency ($\alpha = .84$-$.97$) and convergent validity ($rs = .35$-$.83$).

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Strickland et al., 2013	Interview Skills Rating Instrument. 2 subscales: Response Content and Response Delivery Assessed by standard and situational interview questions. Social Responsiveness Scale (SRS)	Interview Skills Rating Instrument - assessed Response and Delivery Content: Response Content: A 10 item scale measuring content of the participant's responses to 10 interview questions. Response Delivery: 20 items measuring behaviors related to greetings and farewells (handshakes, eye contact, verbal greeting, verbal expression of appreciation at end of interview), as well as the non-verbal behaviors (e.g., body positioning, facial expressions) that accompany verbal responses. 13 items scored on Likert-type scale from 1 (poor) to 4 (excellent) for each response, 7 items scored operationally as 1 (did not occur) or 4 (occurred)	Mean difference scores from all raters between Time 1 and Time 2: treatment group showed a significant positive change at the second interview on the Content Scale (M = .448, SD = .341; Control: M =-.034, SD = .17, F(1,20) = 17.46, p < .000, eta squared = .47), and a trend toward a positive change on the Delivery Scale (Treatment: M = .334 (SD = .229); Control: M = .0252 (SD = .463); F(1,20) = 3.93, p = .062, eta squared = .16). None of the SRS scores correlated significantly with outcomes on the Rating Scales	3 independent raters and in-person interviewer, blind to condition of ppts.	Interview Rating Scale reliability - Significant correlations between all four raters: for Content pearson's r = .71 (p < .01) and for Delivery r = .78 (p < .01).

Paper	Outcome measure names	Outcome definition	Results	Person measuring	Reported validity and reliability
Yang et al., 2017	The Advanced Clinical Solutions for WAIS-IV and WMS-IV Social Perception Subtest (ASC-SP) (Kandalaft et al., 2012; Pearson, 2009); and The Social Attribution Task (SAT) (Triangles task) (Abell, Happe, Frith, 2000)	ASC-SP - to measure emotion recognition abilities through 3 subtests: (a) affect naming (measure of face emotion recognition); (b) Prosody (measure of vocal affect recognition); (c) Pairs (measure of non-literal language interpretation). SAT - to measure a person's theory of mind and abilities to understand others' intentions.	VR-SCT significantly improved emotion recognition - ACS-SP scaled scores from pretreatment (M = 11.41, SD = 4.42) to posttreatment (M = 12.94, SD = 3.51), D = 1.53, S.D. of D = 2.72, t(16) = 2.32, p= 0.03 (two-tailed), Cohen's drm (Lakens, 2013) = 0.37. VR-SCT marginally improved theory of mind - total scores from Triangles task from pretreatment (M = 19.41, SD = 3.89) to posttreatment (M = 20.35, SD = 3.84), D = 0.94, SD of D = 1.95, t(16) = 1.99, p = 0.06 (two-tailed) Cohen's drm (Lakens, 2013) = 0.24, significant one-tailed, p = 0.03.	ASC-SP - administered by trained research staff (not coach); SAT narratives recorded, transcribed and double rated by blind raters.	ACS-SP scores average internal consistency r = 0.69–0.81, test–retest stability coefficient corrected r = 0.60 –0.70, and inter-scorer agreement from 0.98 to 0.99. Triangles SAT test: high test – retest reliability of r= 0.76–0.88 and concurrent validity r =0.78–0.93 [Hu et al., 2010].

4. DISCUSSION

4.1. Aims

The aims of this systematic review for thesis purposes was to examine the feasibility and utility of VR-based interventions for individuals with ASD, with consideration for efficacy and effectiveness of these interventions and any reported real life impact.

Through a systematic search of three databases with key words of interest, a total of 1330 studies were returned. After removing duplicate papers, screening for eligibility, and excluding studies duplicating behavioural data from previous studies within the same research team, a total of 24 articles were considered eligible for the systematic review.

4.2 Feasibility

Newbutt et al., (2016) conducted a study to examine the feasibility of HMD device use in interventions for individuals with ASD. HMD feasibility was queried as a suitable intervention aid due to sensory sensitivities experienced by some ASD individuals (DSM-V, APA, 2013). The Newbutt study demonstrated that the participants could experience enjoyment and a high sense of presence with the immersive technology without adverse side-effects and with low anxiety both prior to and after HMD use. The Newbutt study therefore provided some evidence in the first instance of the feasibility of using such technology. Only two studies in the present review used HMD technology. Several of the studies included in the review report that they were feasibility trials in the first instance, to determine whether the technology and/or intervention approach was going to be amenable to autistic spectrum disorder participants. However, none of the included studies directly measured any potential adverse effects of the technology (in any of its various forms) on their participants. Given potential for adverse effects of nausea, headaches, ocular-motor disturbances and disorientation caused by VR technology (Kennedy, Lane, Berbaum & Lilienthal, 1993; Balk, Bertola & Inman, 2013), the Simulated Sickness Questionnaire (SSQ, Kennedy et al., 1993) was developed and continues to be the most popular method to assess such events in different VR environments. None of the included studies used this questionnaire or any other formal measure of adverse events in VR for the participants and as such, none of the included studies scored positively on the respective question (no. 8) on the Downs and Black Quality Assessment. Future research would benefit from including an assessment of potential adverse events of VR exposure to (i) improve the quality of the study, (ii) to ensure the safety and comfort of participants, and (iii) to bolster support for the feasibility and suitability of VR-based treatment approaches. Nevertheless, whilst the research included

does not explicitly examine adverse effects of the technology, the feasibility of VR use in its variety of forms is supported by the low attrition rates across the studies (only two studies report loss of one participant each), as well as good compliance with the interventions provided, as assessed by the Quality Assessment tool (questions 9 and 20 respectively on the Downs and Black Quality Assessment, 1998).

4.3 Utility

When considering the utility of VR-based interventions for ASD, it is necessary to consider the benefits that such intervention can afford participants, in particular the added benefits over non-VR based interventions. Whilst part of this systematic review aimed to determine the potential added value of VR, this was difficult to determine as only one study directly compared VR-based intervention to a non-VR intervention (Cheng et al., 2018). In this study, VR-based Social Stories and facilitated role-play were provided as the experimental intervention in VR compared to a control intervention of paper-based Social Stories, with the aim to examine effects on complex emotion recognition. Whilst both interventions yielded improvements for the participants, the study demonstrated positive effects of the VR intervention over and above the paper-based intervention. Nevertheless, the article did not provide sufficient details about the comparator intervention to determine the components of the intervention and in which respects the two interventions differed apart from the use of VR to present the Social Stories. As such, it is hard to evaluate from this one study whether the greater improvement was due to VR or other confounding factors. Therefore, with respect to addressing the utility of VR intervention for ASD individuals compared to non-VR intervention, this systematic review highlights the need for future research to assess the effects of VR over non-VR interventions by using appropriate comparator interventions.

If considering utility in its broader sense, not in comparison to other types of intervention, this systematic review goes some way to evaluate this. Five studies were randomised controlled designs comparing VR intervention against a waitlist control (Cheng et al., 2018; Ip et al., 2018; Strickland et al., 2013), or treatment-as-usual (Smith et al., 2015, 2014). Each of the controlled studies demonstrated significant improvement in their respective target skills for participants following VR intervention compared to no intervention. Additionally, these studies performed relatively highly on the Quality Assessment (scores ranging from 63% to 79%), lending greater confidence in the study outcomes and support for the utility of VR intervention in their respective areas of emotion recognition and job interviewing.

Nine studies employed a single-arm design reporting pre-post measures or change over time to assess the effects of VR intervention for ASD individuals (Amaral et al., 2018; Didehbani et al., 2016; Kandalaft et al., 2013; Ke & Im, 2013; Lorenzo et al., 2013; Mitchell et al., 2007; Serret et al., 2014; Stichter et al., 2014; Yang et al., 2017). Similarly, the single-arm studies provided positive support for the utility of VR intervention for the most part when comparing skills pre- to post-intervention. However, quality of research and/or reporting varied greatly across studies (scores ranging from 48% to 75% on the Quality Assessment). Other studies that employed a multiple-probe design or case study designs performed relatively poorly on the Quality Assessment (scores ranging from 32% - 44%, and 31% - 52% respectively) and as such do not provide reliable evidence for the utility of VR-based intervention for ASD individuals.

Of those studied that performed well on the Quality Assessment, the systematic review yielded research supporting the efficacy of VR-based intervention for job interviewing skills (Burke et al., 2018; M. J. Smith et al., 2015, 2014; D. C. Strickland et al., 2013); and the effectiveness of VR for affect recognition (Cheng et al., 2018; Didehbani et al., 2016; Kandalaft et al., 2013; Serret et al., 2014; Yang et al., 2017) and emotional regulation (Ip et al., 2018; Maskey et al., 2014). Additionally, there is some limited support for the effectiveness of VR for developing socio-cognitive reasoning skills (understanding social norms and implementing them for social decision making) (Mitchell et al., 2007); and limited support for general social communication and interaction skills (Ip et al., 2018; Stichter et al., 2014). Though several further studies explored the use of VR for this latter target area, very few studies were of sufficient quality to provide a confident evaluation of the effects of VR intervention. At present there is also insufficient or lacking evidence for the utility of VR intervention for the development of joint attention skills or executive functioning.

4.4 Potential of VR for conversation and interaction skills

Despite the limited research for the beneficial effects of VR intervention on social communication and interaction skills, and the need for better quality research in the area, it is an area of real promise for VR intervention. One of the core deficits of ASD is reciprocal social communication and interaction, which includes problems with initiating and ending conversation, turn-taking in conversation, social approach and responding to social interactions (DSM-V, APA, 2013). The impact on daily life of experiencing difficulties in these areas is enormous, and of particular concern as individuals mature into adolescence and adulthood when family support may be reduced and more independent living is desired by both families and ASD individuals alike. VR intervention, through the form of VR-SCT for example, offers

opportunities to practise and develop skills for conversational exchange with avatars, practise particular social situations to gain feedback, repeat situations and reduce the unpredictability of some social scenarios, and consolidate skills to then take these into real-life situations.

In two studies with children, Stichter et al., (2014) and Ip et al., (2018) used somewhat differing approaches from one another, but similarly provided opportunities for facilitated practice of conversation skills and interaction in social scenarios; Stichter employing a Social Competence Intervention curriculum (a combined approach of CBT, ABA and scaffolded learning); and Ip and colleagues employing interaction guidance through social scenarios with relaxation strategies within a group learning context. Results varied with Stichter reporting positive significant improvement on Social Responsiveness Scale by parent report, but not by teachers. Ip et al., (2018), however, reported pre- to post- significant improvement in social interaction skills, as assessed by the PEP-3. The Ip study was more highly powered with 36 children in each of the treatment and waitlist control groups.

Whilst the Kandalaft et al., 2013, Didehbani et al., 2016 and Yang et al., 2017 studies provided experiences for social conversation and interaction practice through VR-SCT, only the Kandalaft study formally measured effects on conversation skills (the other two measured emotion recognition, social attribution, and executive functioning). In the Kandalaft study et al., (2013) study, eight young adults with high functioning ASD between the ages of 18 and 26 years undertook a 10-session VR-SCT intervention. The VE included various social environments (e.g. a restaurant, coffee house, park, and office building) intended to provide realistic opportunities for the participants to practise and gain feedback on engagement in meaningful social daily-life scenarios with specific objectives relating to, for example, meeting new people and social introductions, initiating conversation with others, celebrating with a friend, and negotiating social decisions. As well as assessing emotion recognition, they assessed conversation skills and participants additionally completed a self-report questionnaire to assess impact on daily functioning following the VR-SCT. The pre-post tests of conversational skills demonstrated a trend in improvement, though not statistically significant. Though it is possible that the study was not sufficiently powered to detect a change in this area.

Studies examining VR intervention for conversational exchange skills and interaction are few and far between, and those that have been conducted are too small in sample size or of insufficient quality to demonstrate efficacy of intervention. However, the studies of quality do demonstrate improvement trends. VR has thus shown promise in aiding individuals to develop the important skills of conversational interchange, as well as its utility for more general social

communication and interaction development in offering opportunities to practise, repeat, learn from others and gain feedback, without unpleasant social consequences.

4.5 Impact in real-life

In the Kandalaft et al., (2013) study reported above, in addition to performance measures, self-reported outcomes from ASD participants were examined. Participants reported positive social benefits as a direct consequence of the intervention. Individuals described how the intervention had developed their overall confidence in social situations, and in turn had impacted on their willingness to participate in social opportunities in everyday life. Similarly, Parsons et al. (2016) report that their ASD participants expressed how they had generalised learning from the VR experience into real life. In their study, participants were required to choose an appropriate place to sit in a café and on a bus, asking members of the public if a seat is available or asking if they could move their belongings to sit down. Participants reported how they had used other public transport and used the skills they had practised in VR, additionally expressing they felt proud for doing so. Unfortunately, the study involved only two participants. Further real-life follow-up information would have been interesting to collect from more participants in this study. Similarly, real-life follow-up data would have been interesting from the Mitchell et al. (2007) study, which employed the same intervention and social scenario demands as the Parsons study. Mitchell and colleagues demonstrated improvements in understanding social norms and implementing them for social decision-making in the VRE, but unfortunately, no real-life follow-up information was gathered from participants in this study.

Although studies such as that by Kandalaft may not have been able to provide statistical significance with respect to results from performance measures of social interaction, it is plausible that the greatest impact of the interventions was in individuals' confidence in participating in social situations and interactions, and therefore widening their life experiences and independent living.

To make solid claims in this area, follow-up data in real life settings is essential, and evaluation of confidence levels and / or anxiety levels would also be useful. Unfortunately, the number of studies that collected this information in the included studies in this review are very few. Two studies evaluated anxiety specifically (Amaral et al., 2018; Maskey et al., 2014), and two evaluated confidence levels in the situations practised (Maskey et al., 2014; Smith et al., 2014). Ten studies conducted follow-up surveys or evaluations with respect to managing situations in real life (Burke et al., 2018; Cheng, Chiang, Ye, & Cheng, 2010; Herrera et al., 2008; Kandalaft et

al., 2013; Lorenzo, Lledó, Pomares, & Roig, 2016; Lorenzo, Pomares, & Lledó, 2013; Maskey et al., 2014; Parsons, Leonard, & Mitchell, 2006; Smith et al., 2015, 2014).

Lorenzo et al., 2013 and Lorenzo et al., 2016 both employed a semi-cave immersive VR environment with children for role-play of social scenarios. Analyses of frequencies of pro-social behaviours, including emotion recognition and appropriate social responsiveness demonstrated improvement within VRE across sessions. These target behaviours were also assessed by the children's teachers in real classroom settings and similarly showed improvements during the period of intervention. Unfortunately, both studies conducted by Lorenzo and colleagues grouped different target behaviours together under appropriate emotional and behavioural responses. As such, it is difficult to determine which specific social skills improved or increased in frequency and which decreased. However, the studies demonstrated generalisation of VR effects to the real-life classroom. Data on effects post-intervention period were not gathered and it is thus not possible to state whether any positive improvements were maintained. In contrast, Cheng et al. (2010) reported some parents' observations on the three children who undertook a CVLE empathy programme. Parents provided reports of their children demonstrating verbal expressions of empathy and care towards others, thus carrying over social behaviours from VR into their daily life.

In adult research, Burke et al. (2018) and Smith et al., (2014, 2015) conducted pre- and post-evaluations of VR job interviewing skills interventions with real-life follow-up interviews. Burke compared performance in several skills from baseline to final session and with a final face-to face real-life interview. Although there was a decrease in performance scores between the last VR session and real-life interview, the results demonstrate a measurable and significant improvement in interviewing skills across sessions which are translated into real-world experience. Smith et al., (2014) similarly assessed effects of VR job interviewing intervention (VR-JIT) with real-life pre-post interviews and compared VR-JIT intervention to a TAU group. They found total scores showed a significant large effect for the intervention compared to TAU group. Additionally, self-confidence measures demonstrated improvement with a large effect size for the intervention group. Analyses of their 6 month follow-up survey of this intervention (Smith et al., 2015) demonstrated that the pre-post VR-JIT improvement correlated with completing more interviews for a competitive position in real life, and that VR-JIT trainees had almost eight times greater odds of accepting an offer for a position.

Maskey et al. (2014) conducted a study that evaluated VR intervention on anxiety, confidence in particular settings and also conducted a follow-up. Their study with children with ASD

employed a four-sided cave immersive VR environment to tackle children's specific fears and phobias over four sessions, using a CBT approach. Treatment included emotion identification, techniques in relaxation, and using positive thoughts to cope when experiencing anxiety. Specific VR scenes were programmed and individualised to each participant's specific phobia (these included crossing busy roads, pigeon phobia, shopping and speaking to people in a shop, alighting a crowded bus, being a passenger in a car, crossing a bridge with water underneath, and speaking in class). The scene always began at a challenge level which the child rated as low anxiety-provoking, and the scene was gradually increased in challenge over the four sessions. For example, for a child afraid of shopping, the scene began with simply going into an empty shop and taking something to the counter. Gradually, the number of people in the shop increased, or the length of the verbal exchange with the shop assistant was increased. This was conducted at a rate that enabled the participant to feel as comfortable and relaxed as possible. The scene could be repeated as many times as required by the child, the therapist would check how the child rated their anxiety, and used relaxation techniques and thought challenges. The parents were able to watch the therapist interacting with their child and learn the CBT techniques being used. Confidence ratings from the children and their parents were also gathered with respect to tackling their target situation. These demonstrated that seven out of the nine children increased their confidence from pre-to final VR session, which was maintained at two weeks post intervention. Anxiety was rated by both parents and the child at four time-points pre-treatment, 6 weeks post-treatment, 6 months, and 12-16 months post-treatment to determine response to treatment. Eight out of the nine children were deemed treatment respondents according to the target behaviour ratings, and seven of the children had anxiety ratings within the normal range at 12-16 months follow-up. In summary, eight of the children experienced benefit from the VR intervention within six weeks, going on to face their phobia in real life. Four of the children overcame their fear within this time. At 6 months and 12–16 months follow-up, improvement in target behaviours was maintained or improved for all children. Parents reported that their children's management of anxiety generalised to other areas of life and activities.

All of the above studies therefore demonstrate how the effects of VR intervention can be generalised to real-life, and in some cases generalised across contexts not specific to the VR environment scenario experienced. This said, insufficient numbers of studies report on the effectiveness of intervention outside of the controlled environment, and few report on it sufficiently well.

4.6 Clinical theory considerations

However VR intervention is delivered, be it with HMD, a four-sided cave or on desktop, the intervention content should be informed by clinical theory with an evidence base for its efficacy or effectiveness. Most studies included in this systematic review have been exploratory in nature, with small sample sizes and provide first steps in developing VR intervention. As such, researchers of these studies have not been limited by evidence-based intervention protocol. However, it is potentially for this reason that more studies have not reached significance levels of improvement from pre- to post-intervention. If such intervention is to be taken to the next level to offer it to clinical populations for healthcare intervention, it is essential, for ethical reasons, that intervention be informed by clinical evidence.

With respect to intervention without VR, Social Skills Training or Social Cognition Training is the most widely used psychosocial intervention method to address social communication and interaction difficulties in ASD. There is an increasing body of research to evidence its effectiveness for developing social interaction skills (Gevers, Clifford, Mager, & Boer, 2006; Laugeson, Frankel, Gantman, Dillon, & Mogil, 2012; Laugeson et al., 2009; Turner-Brown, Perry, Dichter, Bodfish, & Penn, 2008; Williams White, Keonig, & Scahill, 2007). However, few studies in the present systematic review employed an SCT approach. Those that did (Didehbani et al., 2016; Kandalaft et al., 2013; Yang et al., 2017), provide some evidence towards its effectiveness for emotion recognition improvement, but those that explored conversation and communications skills were of insufficient size or quality to clearly determine the benefits of VR-SCT in this area. However, some self-reports from participants suggest at the promise of its effectiveness in aiding individuals expand their social life experiences.

With respect to psychotherapy approaches for ASD, CBT is the predominant model, with evidence for its utility for social cognition and interaction difficulties (Bauminger, 2002, 2007; Wood et al., 2015), as well as adapted-CBT for anxiety in ASD (Chalfant et al., 2007; Kreslins et al., 2015; J. A. Reaven et al., 2009; J. Reaven & Hepburn, 2003; Sze & Wood, 2007, 2008). Nevertheless, only one study (Maskey et al., 2014) employed a CBT approach in their VR intervention, and two employed elements of CBT (Ip et al., 2018; Stichter et al., 2014). With this in mind, it is important that future research, and clinicians and researchers alike, inform their intervention protocol for a target skill with an evidence-based intervention approach.

4.7 Immersiveness

As Witmer and Singer (1998) note, the greater the immersive experience, the more ‘powerful’ the tool. Increased immersion can influence the degree of ecological validity, and in turn the

extent of skill transfer (Wang & Reid, 2011). It could be argued that HMD and four-sided cave environments provide superior VR experience in terms of immersiveness compared to desktop flatscreen 3D experiences, and in turn are likely to yield greater effects and effectiveness. However, it can also be argued that immersion is a subjective experience, and effects would in part be due to an individual's capacity to block out surrounding sensory input. Some argue that presence can be regarded as a quasi- measure of attention in virtual environments (Waterworth and Waterworth, 2001).

Given the large differences across the studies included in this systematic review, it was not appropriate to conduct a meta-analyses at this stage to determine whether the type of VR technology yields greater effects over others. Furthermore, none of the included studies in this systematic review gathered data on participants' experiences of immersiveness in the different VREs to determine their subjective experience of presence and immersiveness in the VRE. As such it could not determine whether differences in effects are due to VR technology, subjective sense of presence, or specific components of the intervention.

4.8 Clinical Implications

To summarise some points made so far, the evidence from this systematic review in general has some positive results in favour of VR-based interventions for people with ASD, within specific target areas. There is a relatively strong evidence base in favour of VR intervention for developing job interviewing skills, which include skills in conversational interchange and communication of self-advocacy. Some of this evidence demonstrating the benefits of VR intervention for job interviewing comes from RCTs (Strickland et al., 2013; Smith et al., 2015, 2014), and therefore give further weight to the evidence for its utility. Further evidence is provided for the utility of VR intervention for the purposes of developing affect recognition, again with some from an RCT (Cheng et al., 2018). The use of VR intervention for emotional regulation skills also has some strong support from an RCT (Ip et al., 2018) as well as some good quality non-RCT research (Maskey et al., 2014).

Clinicians and service developers should not be afraid to start employing such technology-based interventions in their clinics, and in fact some projects are currently underway to provide such VR-based services, e.g. The Northumberland Tyne and Wear NHS Foundation Trust in collaboration with Newcastle University and Third Eye NeuroTech, (www.passionprojectfoundation.org.uk/thirdeye). However, when contemplating implementing such intervention clinically, the evidence must be interpreted cautiously and

understood within the terms of the limitations of the research conducted to date. Several factors must therefore be taken into account and given careful consideration: It is yet undetermined the extent of effectiveness (and cost effectiveness) of VR-intervention compared to traditional face-to-face intervention. Similarly, with respect to comparison studies, it is yet undetermined which type of VR intervention technology (HMDs, CAVE environments, desktop delivery, etc.) would yield the most clinically significant (and simultaneously cost effective) results. Additionally, only very limited information is available on the long-term impact in the real-world for recipients of VR intervention. Interventions should also draw upon theoretical approaches, intervention components and clinical protocols that already have an evidence base without VR delivery. Finally, intervention development should consider the service user's opinion on the approach, and most importantly, what the ASD client deems is the most salient challenge for them as an individual to determine the aim of the intervention.

4.9. Limitations

4.9.1 Applicability of findings

On the Quality Assessment, few studies gained a score on the questions relating to whether the selection of participants was representative of the population (questions 11 and 12). This is due to the heterogeneity of the ASD population in general and the fact that most studies in this systematic review recruited only high functioning ASD individuals (IQ > 70). In fact, only one included study recruited participants with and without learning disability. Studies have been conducted examining VR-based intervention with autistic individuals with learning disabilities (e.g. Josman, Tamar Weiss, Ben-Chaim, & Friedrich, 2008; Strickland, Mcallister, Coles, & Osborne, 2007). However, these studies were generally excluded from this review as they did not meet the inclusion criteria of target intervention being for core deficits of ASD or anxiety. Rather, these studies targeted other skills, such as road crossing and fire safety (Strickland, Mcallister, Coles, & Osborne, 2007). This current systematic review therefore provides data on the utility of VR intervention for a subset of the ASD population, specifically, people considered high functioning/without a learning disability. The studies included in the present review involve interventions that require verbal and/ or reading capacities. As such, the applicability of the findings from this systematic review are limited to this subset of the general population of people with ASD. Different VR interventions with distinct targets will be suitable for different populations within the general ASD population. For individuals with ASD and learning disability, it is possible that areas of difficulty other than social communication may be considered more

pressing for intervention target purposes. Further research and/or another systematic review of evidence would be required to determine the target areas that would benefit from VR-intervention for individuals with ASD with learning disabilities and greater support needs.

4.9.2 Quality Assessment

Use of a quality assessment tool was important to assess risk of bias in the studies included in this review. The Downs and Black (1998) tool was chosen due to its potential to evaluate both randomised and non-randomised studies. However, a lack of a manual or sufficient guidance on interpreting the questions within the tool made it difficult to assess non-randomised studies due to the wording of some questions. As such, the Downs and Black assessment tool was adapted to allow a response of 'not applicable' to some studies and a percentage of total applicable scores given to studies. This adapted version of the quality assessment has not been validated. Sterne et al. (2016) similarly point out that the lack of guidance can cause different users to interpret instructions differently. As such it would have been useful to have one or more independent raters assess all the studies using the tool, rather than a small sample of them.

4.10 Future Research

The current systematic review has highlighted several areas of focus for future studies to improve the quality of research in the area of VR intervention in ASD, as well as to provide greater clarification on the relative effectiveness/efficacy of the tech-based intervention, and its impact in real life terms. Firstly, when beginning this systematic review, it was anticipated that more studies would have explored interventions for anxiety with VR, or if not a primary target of intervention, would have used outcome measures to examine anxiety levels in given situations. However, the systematic review yielded only two studies that evaluated anxiety specifically (Amaral et al., 2018; Maskey et al., 2014), and two that examined confidence levels (Maskey et al., 2014; Smith et al., 2014). Prevalence rates of anxiety in the ASD population are high, and anxiety can have debilitating effects on individuals in their day-to-day life. The convergence of ASD traits with anxiety can prevent a person from engaging in social situations, continuing education, holding down a job and generally participating in the life they wish to lead. This in turn has implications for an individual's quality of life, potentially leading to social isolation, feelings of lack of purpose, and depression. A person's difficulties can grow exponentially if anxiety is not addressed. Further studies would therefore be useful to examine effects of VR intervention for anxiety management specifically, preferably with larger sample sizes than those conducted to date. Alternatively, if the target of intervention is less specifically anxiety and rather conversational skills or social reasoning for instance, the research would

benefit from an evaluation of the impact on anxiety and confidence levels following skill development in these areas. As such, future research should consider employing outcome measures of anxiety and self-confidence levels, suitably adapted to individuals with ASD (as discussed in Section 1.2).

On a similar line, it is important that research consider the impact of VR-intervention outside of the VR environment; that is, studies would benefit from systematically evaluating the effects of VR training not simply pre- and post-intervention but also at follow-up intervals for scenarios and situations in real life that the VR environment emulated during intervention. As the ultimate purpose of VR intervention is to assist individuals to engage in daily life, it is surprising that such few studies in this systematic review evaluated intervention impact in the real world. In addition to outcome measures for particular skill development being employed for evaluation in the long-term, it would be particularly interesting to gather data from participating individuals on the impact on their quality of life. Even if objective outcome measures of skill development do not show significant effects pre-post intervention, it is plausible that the ability to practice and receive feedback on socially demanding scenarios that VR affords, can have a significant impact on confidence levels, an individual's ability to face and engage in social scenarios, and in turn improve their quality of life.

This systematic review has highlighted that whilst VR intervention may indeed afford benefits for individuals in specific areas, the added-value of VR intervention compared to more traditional interventions cannot as yet be determined. It is clear that VR offers many possibilities and opportunities beyond those of traditional interventions for people with ASD (as outlined in section 1.5.2.1), but it is unclear whether these possibilities equate to clinical improvements and effectiveness above and beyond traditional non-VR methods. Only one study included in this review directly compared effects of VR-based intervention to a non-VR-based intervention (Cheng et al., 2018) and reporting of the components of the traditional method was poor. As such, research has yet to determine the relative effectiveness, and cost effectiveness, of VR training compared to intervention delivered without VR. Future research would benefit from exploring this, especially as it potentially has implications for the treatment and training offered in clinical settings. It appears somewhat very limited to only compare target skills pre- to post-intervention. It is crucial to establish a suitable comparator intervention against which to compare effects in order to identify potential added-value and determine whether VR intervention is worthy of the financial and personal investment required. In a similar vein, moving forward, it is important to determine which VR delivery systems (e.g. HMD, desktop,

CAVE environment, etc.) provide the most clinically significant results. Whilst it may be argued that systems that offer the most immersive experience will offer the most gains, immersiveness can be a subjective experience. As some VR delivery systems are more expensive than others and may require customised rooms for their implementation, it is important that research examine whether the same interventive approach may yield different effects when delivered through different VR systems. This is essential to determine whether the effects of a particular delivery system warrant any additional costs associated with 'more immersive' experiences compared to other delivery systems. As part of this process, studies should additionally consider evaluation of any adverse effects of these different delivery systems.

Finally, where possible, research participants with ASD should be consulted as part of the research process and the development of intervention. Ideally, research would aim to adopt a neurodiversity model in the development of a study, whereby the target of intervention/training is not determined exclusively by non-autistic people, but instead incorporates the views, opinions and insight from those who are to receive the intervention. The benefits of any given intervention are only as important as the person in receipt perceives them. As such, capturing the thoughts and experiences of individuals with ASD about areas of life they find most challenging is essential to develop relevant intervention programmes, which ultimately aim to improve quality of life. By doing so, the research process itself becomes the first step to empower individuals with ASD, and to foster greater respect for and value of individuals with diverse neural processing.

5. CONCLUSIONS

Twenty-four studies were included in this systematic review of interventions with the use of virtual reality for participants with ASD. The studies varied considerably in terms of focus of intervention, components of intervention, and type of VR system. Quality of studies also varied greatly in terms of risk of bias. Many studies had small sample sizes to only allow data inspection at the individual level, or were not sufficiently powered to detect a significant change pre- to post-treatment in some target areas, making it difficult to generalise findings from these studies to the general ASD population. This said, the systematic review yielded five randomised controlled studies and some single-arm designed studies which performed well on the quality assessment and provide evidence for the efficacy and effectiveness of VR intervention in the improvement of affect recognition and emotion regulation skills, as well as job interviewing

skills. An area of particular promise for VR intervention is its benefit in developing understanding and competences in social interactional and conversational skills. Nevertheless, there is a need for further rigorous empirical testing in this area. In particular, it is imperative going forward that VR intervention research be driven by clinical theory and evidence-based approaches. This can mean considering approaches that are non-CBT based, such as Acceptance and Commitment Therapy, which has less emphasis on cognitive restructuring and may be more, or equally, suitable compared to ASD-adapted CBT for ASD individuals. Future research could benefit the evidence base for VR intervention if it were to empirically evaluate VR-based treatments compared to non-VR based treatments. This would allow formal assessment of the perceived added-benefits that VR offers over traditional interventions and determine whether VR features ameliorate treatment effects. At present, it would seem that VR allows a more controlled means of graded exposure to particular environments and offers an alternative to imaginal exposure which ASD individuals might find difficult. Additionally, VR offers an environment in which to practise as many times as a person might need without adverse consequences. Some studies note that effects may be attributable to the practice effects of repetitive use within particular VR scenarios, rather than the VR-based protocol itself. However, this has a positive connotation in that this is exactly what VR affords that real-life intervention cannot, i.e. the opportunity to repeat, repeat and repeat, to practise, gain confidence, learn from feedback and develop skills in an environment with no consequences. Some studies have provided user feedback on these experiences and the impact in their daily life, and the anecdotal reports offer insight into how VR has potential to build confidence and reduce anxieties to participate in social situations and interact with others. This in turn, can have important consequences in terms of widening life experiences, reducing isolation and improving quality of life. To make the biggest impact in this area, it is important that individuals with ASD are themselves consulted about what they would find most useful from a VR-based intervention opportunity, rather than intervention targets being directed by researchers, clinicians or carers. It could be argued that some current interventions are designed to enable individuals with ASD to appear more neurotypical in their behaviours. The greatest achievement would be to embrace individuals' neurodiversity, and to develop an intervention that enables individuals to navigate and manage social situations without autistic traits or concomitant anxiety inhibiting them, to enable them to engage in the life they want to the fullest.

REFERENCES

Amaral, C., Mouga, S., Simões, M., Pereira, H. C., Bernardino, I., Quental, H., … Oliveira, G. (2018). A Feasibility Clinical Trial to Improve Social Attention in Autistic Spectrum Disorder (ASD) Using a Brain Computer Interface, *12*(July), 1–13. https://doi.org/10.3389/fnins.2018.00477

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th edition). Washington, DC.

Anderson, S., & Morris, J. (2006). Cognitive behaviour therapy for people with Asperger syndrome. *Behavioural and Cognitive Psychotherapy*, *34*(3), 293–303. https://doi.org/10.1017/S1352465805002651

Armstrong, T. (2011). The Power of Neurodiversity: Unleashing the Advantages of Your Differently Wired Brain (published in hardcover as Neurodiversity). Da Capo Lifelong Books.

Attwood, T. (2004). Cognitive Behaviour Therapy for Children and Adults with Asperger ' s Syndrome. *Behavior Change*, *21*(3), 147–161.

Austin, D. W., Abbott, J. M., & Carbis, C. (2008). The Use of Virtual Reality Hypnosis with Two Cases of AutismSpectrum Disorder: A Feasibility Study. *Contemporary Hypnosis*, *25*(2), 102–109. https://doi.org/10.1002/ch

Balk, S. A., Bertola, M. A., & Inman, V. W. (2013). Simulator sickness questionnaire: Twenty years later.

Bauminger, N. (2002). The Facilitation of Social-Emotional Understanding and Social Interaction in High-Functioning Children with Autism : Intervention Outcomes, *32*(4).

Bauminger, N. (2007). Brief Report: Individual Social-Multi-Modal Intervention for HFASD. *Journal of Autism and Developmental Disorders*, *37*, 1593–1604. https://doi.org/10.1007/s10803-006-0245-4

Bauminger, N., Gal, E., Goren-Bar, D., Kupersmitt, J., Pianesi, F., Stock, O., … Zancanaro, M. (2007). Social Intelligence Design 2007. In *Social Intelligence Design 2007 CTIT* (pp. 15–25). Retrieved from http://eprints.eemcs.utwente.nl/10263/01/sid07.pdf

Bellini, S., Peters, J.K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. Remed Spec Ed. 28:153-162.

Biocca, F., & Levy, M. R. (1995). Virtual reality as a communication system. Communication in the age of virtual reality, 15-31.

Bradley, E. a, Summers, J. a, Wood, H. L., & Bryson, S. E. (2004). Comparing rates of psychiatric and behavior disorders in adolescents and young adults with severe intellectual disability with and without autism. *Journal of Autism and Developmental Disorders*, *34*(2), 151–161. https://doi.org/10.1023/B:JADD.0000022606.97580.19

Broeren, J., Claesson, L., Goude, D., Rydmark, M., & Sunnerhagen, K. S. (2008). Virtual rehabilitation in an activity centre for community-dwelling persons with stroke: The possibilities of 3-dimensional computer games. *Cerebrovascular Diseases*, *26*(3), 289–296. https://doi.org/10.1159/000149576

Burke, S. L., Bresnahan, T., Li, T., Epnere, K., Rizzo, A., Partin, M., … Trimmer, M. (2018). Using Virtual Interactive Training Agents (ViTA) with Adults with Autism and Other Developmental Disabilities. *Journal of Autism and Developmental Disorders*, *48*(3), 905–912. https://doi.org/10.1007/s10803-017-3374-z

Cameirão, M. S., Badia, S. B. I., Oller, E. D., & Verschure, P. F. M. J. (2010). Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: Methodology, design, psychometrics, usability and validation. *Journal of NeuroEngineering and Rehabilitation*, *7*(1), 1–14. https://doi.org/10.1186/1743-0003-7-48

Cath, D. C., Ran, N., Smit, J. H., Van Balkom, A. J. L. M., & Comijs, H. C. (2007). Symptom overlap between autism spectrum disorder, generalized social anxiety disorder and obsessive-compulsive disorder in adults: A preliminary case-controlled study. *Psychopathology*, *41*(2), 101–110. https://doi.org/10.1159/000111555

Chalfant, A. M., Rapee, R., & Carroll, L. (2007). Treating anxiety disorders in children with high functioning autism spectrum disorders: A controlled trial. *Journal of Autism and Developmental Disorders*, *37*(10), 1842–1857. https://doi.org/10.1007/s10803-006-0318-4

Cheng, Y., Chiang, H., Ye, J., & Cheng, L. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education*, *55*, 1449–1458. https://doi.org/10.1016/j.compedu.2010.06.008

Cheng, Y., Huang, C. L., & Yang, C. S. (2015). Using a 3D immersive virtual environment system to enhance social understanding and social skills for children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *30*(4), 222–236. https://doi.org/10.1177/1088357615583473

Cheng, Y., & Huang, R. (2012). Research in Developmental Disabilities Using virtual reality environment to improve joint attention associated with pervasive developmental disorder. *Research in Developmental Disabilities*, *33*(6), 2141–2152.

https://doi.org/10.1016/j.ridd.2012.05.023

Cheng, Y., Luo, S. yun, Lin, H. chieh, & Yang, C. S. (2018). Investigating mobile emotional learning for children with autistic spectrum disorders. *International Journal of Developmental Disabilities*, *64*(1), 25–34. https://doi.org/10.1080/20473869.2016.1206729

Cheng, Y., & Ye, J. (2010). Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment - The pilot study. *Computers and Education*, *54*(4), 1068–1077. https://doi.org/10.1016/j.compedu.2009.10.011

Cobb, S., Beardon, L., Eastgate, R., Glover, T., Kerr, S., Neale, H., … Wilson, J. (2002). Applied virtual environments to support learning of social interaction skills in users with Asperger's Syndrome. *Digital Creativity*, *13*(1), 11–22. https://doi.org/10.1076/digc.13.1.11.3208

Costello, E. J., Egger, H. L., & Angold, A. (2005). The developmental epidemiology of anxiety disorders: Phenomenology, prevalence, and coborbidity. *Child and Adolescent Psychitric Clinics of North America*, *14*, 631–648.

Deeks, J., Dinnes, J., D'Amico, R., Sowden, A., Sakarovitch, C., Song, F., … Altman, D. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, *7*(27).

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13(1)*(20), 1–14. https://doi.org/10.1167/13.1.20.Introduction

Didehbani, N., Allen, T., Kandalaft, M., Krawczyk, D., & Chapman, S. (2016). Virtual Reality Social Cognition Training for children with high functioning autism. *Computers in Human Behavior*, *62*, 703–711. https://doi.org/10.1016/j.chb.2016.04.033

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, *52*(6), 377–384. https://doi.org/10.1136/jech.52.6.377

Farrugia, S., & Hudson, J. (2006). Anxiety in Adolescents With Asperger Syndrome: Negative Thoughts, Behavioral Problems, and Life Interference. *Focus on Autism and Other Developmental Disabilities*, *21*(1), 25–35. https://doi.org/10.1177/10883576060210010401

Ferrer-García, M., & Gutiérrez-Maldonado, J. (2012). The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: A review of the literature. *Body Image*, *9*(1), 1–11.

https://doi.org/10.1016/j.bodyim.2011.10.001

Foley, T., & Woollard, J. (2018). *The digital future of mental healthcare and its workforce*. Retrieved from https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-Mental-health-paper.pdf

Gevers, C., Clifford, P., Mager, M., & Boer, F. (2006). Brief report: A theory-of-mind-based social-cognition training program for school-aged children with pervasive developmental disorders: An open study of its effectiveness. *Journal of Autism and Developmental Disorders*, *36*(4), 567–571. https://doi.org/10.1007/s10803-006-0095-0

Gillott, A., & Standen, P. J. (2007). Levels of anxiety and sources of stress in adults with autism. *Journal of Intellectual Disabilities*, *11*(4), 359–370. https://doi.org/10.1177/1744629507083585

Gjevik, E., Eldevik, S., Fjæran-Granum, T., & Sponheim, E. (2011). Kiddie-SADS reveals high rates of DSM-IV disorders in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(6), 761–769. https://doi.org/10.1007/s10803-010-1095-7

Halabi, O., Abou El-Seoud, S., Alja'am, J., Alpona, H., Al-Hemadi, M., & Al-Hassan, D. (2017). Design of Immersive Virtual Reality System to Improve Communication Skills in Individuals with Autism. *International Journal of Emerging Technologies in Learning (IJET)*, *12*(05), 50. https://doi.org/10.3991/ijet.v12i05.6766

Herrera, G., Alcantud, F., Jordan, R., Blanquer, A., Labajo, G., & De Pablo, C. (2008). Development of symbolic play through the use of virtual reality tools in children with autistic spectrum disorders: Two case studies. *Autism*, *12*(2), 143–157. https://doi.org/10.1177/1362361307086657

Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Insel, T., & Sahakian, B. J. (2012). A plan for mental illness. Nature, 483, 269.

Ip, H. H. S., Wong, S. W. L., Chan, D. F. Y., Byrne, J., Li, C., Yuan, V. S. N., … Wong, J. Y. W. (2018). Enhance emotional and social adaptation skills for children with autism spectrum disorder: A virtual reality enabled approach. *Computers and Education*, *117*, 1–15. https://doi.org/10.1016/j.compedu.2017.09.010

Josman, N., Tamar Weiss, N., Ben-Chaim, H. M., & Friedrich, S. (2008). Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism.

International Journal on Disability and Human Development, *7*(1), 49–56. https://doi.org/10.1515/IJDHD.2008.7.1.49

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *43*, 34–44. https://doi.org/10.1007/s10803-012-1544-6

Ke, F., & Im, T. (2013). Virtual-Reality-Based Social Interaction Training for Children with High-Functioning Autism, *0671*. https://doi.org/10.1080/00220671.2013.832999

Ke, F., & Lee, S. (2016). Virtual reality based collaborative design by children with high-functioning autism : design- based flexibility , identity , and norm construction, *4820*(May 2015). https://doi.org/10.1080/10494820.2015.1040421

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), 203-220.Kerns, Connor M., Moskowitz, L. J., Rosen, T., Drahota, A., Wainer, A., Josephson, A. R., … Lerner, M. D. (2018). A Multisite, Multidisciplinary Delphi Consensus Study Describing “Usual Care” Intervention Strategies for School-Age to Transition-Age Youth With Autism. *Journal of Clinical Child and Adolescent Psychology*, *00*(00), 1–22. https://doi.org/10.1080/15374416.2017.1410826

Kerns, C. M., Kendall, P. C., Berry, L., Souders, M. C., Franklin, M. E., Schultz, R. T., … Herrington, J. (2014). Traditional and Atypical Presentations of Anxiety in Youth with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *44*(11), 2851–2861. https://doi.org/10.5588/ijtld.16.0716.Isoniazid

Koegel, R. L., Koegel, L. K., Vernon, T. W., & Brookman-Frazee, L. I. (2010). Empirically supported pivotal response treatment for children with autism spectrum disorders. Evidence-based psychotherapies for children and adolescents, 2, 327-344.

Kreslins, A., Robertson, A. E., & Melville, C. (2015). The effectiveness of psychosocial interventions for anxiety in children and adolescents with autism spectrum disorder: A systematic review and meta-analysis. *Child and Adolescent Psychiatry and Mental Health*, *9*(1), 1–12. https://doi.org/10.1186/s13034-015-0054-7

Kuriakose, S., & Lahiri, U. (2017). Design of a Physiology-Sensitive VR-Based Social Communication Platform for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(8), 1180–1191.

Kuusikko, S., Pollock-Wurman, R., Jussila, K., Carter, A. S., Mattila, M. L., Ebeling, H., … Moilanen, I. (2008). Social anxiety in high-functioning children and adolescents with

autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *38*(9), 1697–1709. https://doi.org/10.1007/s10803-008-0555-9

Lahiri, U., Warren, Z., & Sarkar, N. (2011). Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism. *IEEE Trans Neural Syst Rehabil Eng*, *19*(4), 1–27. https://doi.org/10.1038/jid.2014.371

Lahiri, Uttama, Bekele, E., Dohrmann, E., Warren, Z., & Sarkar, N. (2015). A Physiologically Informed Virtual Reality Based Social Communication System for Individuals with Autism. *Journal of Autism and Developmental Disorders*, *45*(4), 919–931. https://doi.org/10.1007/s10803-014-2240-5

Lang, R., Regester, A., Lauderdale, S., Ashbaugh, K., & Haring, A. (2010). Treatment of anxiety in autism spectrum disorders using cognitive behaviour therapy: A systematic review. *Developmental Neurorehabilitation*, *13*(1), 53–63. https://doi.org/10.3109/17518420903236288

Lau, H. M., Smit, J. H., Fleming, T. M., & Riper, H. (2017). Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Frontiers in psychiatry, 7, 209.

Laugeson, E. A., Frankel, F., Gantman, A., Dillon, A. R., & Mogil, C. (2012). Evidence-based social skills training for adolescents with autism spectrum disorders: The UCLA PEERS program. *Journal of Autism and Developmental Disorders*, *42*(6), 1025–1036. https://doi.org/10.1007/s10803-011-1339-1

Laugeson, E. A., Frankel, F., Mogil, C., & Dillon, A. R. (2009). Parent-assisted social skills training to improve friendships in teens with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(4), 596–606. https://doi.org/10.1007/s10803-008-0664-5

Lecavalier, L., Wood, J. J., Halladay, A. K., Jones, N. E., Aman, M. G., Cook, E. H., … Scahill, L. (2014). Measuring anxiety as a treatment endpoint in youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(5), 1128–1143. https://doi.org/10.1007/s10803-013-1974-9

Lerner, M. D., White, S. W., & McPartland, J. C. (2012). Mechanisms of change in psychosocial interventions for autism spectrum disorders. *Dialogues in Clinical Neuroscience*, *14*(3), 307–318. https://doi.org/10.1161/01.RES.0000109414.78907.F9

Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., … Lainhart, J. E. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *Journal of Autism and Developmental Disorders*, *36*(7), 849–861. https://doi.org/10.1007/s10803-006-0123-0

Lickel, A., Maclean, W. E., Blakeley-Smith, A., & Hepburn, S. (2012). Assessment of the Prerequisite Skills for Cognitive Behavioral Therapy in Children with and Without Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(6), 992–1000. https://doi.org/10.1111/mec.13536.Application

Liu, X., Wu, Q., Zhao, W., & Luo, X. (2017). Technology-facilitated diagnosis and treatment of individuals with autism spectrum disorder: An engineering perspective. Applied Sciences, 7(10), 1051.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 557–564.

Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers and Education*, *98*, 192–205. https://doi.org/10.1016/j.compedu.2016.03.018

Lorenzo, G., Pomares, J., & Lledó, A. (2013). Inclusion of immersive virtual learning environments and visual control systems to support the learning of students with Asperger syndrome. *Computers & Education*, *62*, 88–101. https://doi.org/10.1016/j.compedu.2012.10.028

Loveland, K. A. (2005). Social-emotional impairment and self-regulation in autism spectrum. Emotional development: Recent research advances, 365.

MacNeil, B. M., Lopes, V. A., & Minnes, P. M. (2009). Anxiety in children and adolescents with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *3*(1), 1–21. https://doi.org/10.1016/j.rasd.2008.06.001

Maskey, M., Lowry, J., Rodgers, J., McConachie, H., & Parr, J. R. (2014). Reducing specific phobia/fear in young people with autism spectrum disorders (ASDs) through a virtual reality environment intervention. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0100374

Matson, J. L. (1995). The Diagnostic Assessment for the Severely Handicapped-Revised (DASH-II). Baton Rouge, LA: Scientific Publishers.

Matson, J. L., & Nebel-Schwalm, M. S. (2007). Comorbid psychopathology with autism spectrum disorder in children: An overview. *Research in Developmental Disabilities*, *28*(4), 341–352. https://doi.org/10.1016/j.ridd.2005.12.004

Mayes, S. D., Calhoun, S. L., Aggarwal, R., Baker, C., Mathapati, S., Molitoris, S., & Mayes, R. D. (2013). Unusual fears in children with autism. *Research in Autism Spectrum Disorders*,

7(1), 151–158. https://doi.org/10.1016/j.rasd.2012.08.002

Mesa-Gresa, P., Gil-Gómez, H., Lozano-Quilis, J. A., & Gil-Gómez, J. A. (2018). Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: an evidence-based systematic review. Sensors, 18(8), 2486.

Milton, D. (2014). So what exactly are autism interventions intervening with? *Good Autism Practice*, 15 (2), 6-14.

Mishkind, M. C.; Norr, A. M.; Katz, A. C.; Reger, G. M. (2017) Review of Virtual Reality Treatment in Psychiatry: Evidence Versus Current Diffusion and Use. Current Psychiatry Rep. 19, 80.

Mitchell, P., Parsons, S., & Leonard, A. (2007). Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(3), 589–600. https://doi.org/10.1007/s10803-006-0189-8

Moher, D, Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … LA., Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.*, *4*(1), 1–19.

Moher, David, Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Moore, D., Cheng, Y., McGrath, P., & Fan, Y. F. Y. (2005). Collaborative virtual environment technology for people with autism. *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, *20*(4), 231–243. https://doi.org/10.1109/ICALT.2005.85

Muris, P., Steerneman, P., Merckelbach, H., Holdrinet, I., & Meesters, C. (1998). Comorbid anxiety symptoms in children with pervasive developmental disorders. *Journal of Anxiety Disorders*, *12*(4), 387–393. https://doi.org/10.1016/S0887-6185(98)00022-X

Newbutt, N., Sung, C., Kuo, H. J., & Leahy, M. J. (2016). The potential of virtual reality technologies to support people with an autism condition: A case study of acceptance, presence and negative effects. *Annual Review of Cybertherapy and Telemedicine*, 149–154.

Norcross, J. C., Hedges, M., & Prochaska, J. O. (2002). The Face of 2010: A Delphi Poll on the future of psychotherapy. *Professional Psychology: Research and Practice*, *33*, 316–322. https://doi.org/10.1037//0735-7028.13.5.620

Ollendick, T. H., & White, S. (2012). The presentation and classification of anxiety in autism spectrum disorder. *Clinical Psychology: Science and Practice*, *19*(4), 323–347. https://doi.org/10.1111/cpsp.12009

O'Riordan MA, Plaisted KC, Driver J, Baron-Cohen S (2001) Superior visual search in autism. Journal of Experimental Psychology: Human Perception and Performance 27: 719

Parsons, S, & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disability Research Volume*, *46*, 430–443.

Parsons, Sarah. (2015). Learning to work together: Designing a multi-user virtual reality game for social collaboration and perspective-taking for children with autism. *International Journal of Child-Computer Interaction*, *6*, 28–38. https://doi.org/10.1016/j.ijcci.2015.12.002

Parsons, Sarah, Leonard, A., & Mitchell, P. (2006). Virtual environments for social skills training : comments from two adolescents with autistic spectrum disorder, *47*, 186–206. https://doi.org/10.1016/j.compedu.2004.10.003

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., … Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. https://doi.org/10.1001/archderm.1985.01660090059014

Reaven, J. A., Blakeley-Smith, A., Nichols, S., Dasari, M., Flanigan, E., & Hepburn, S. (2009). Cognitive-behavioral group treatment for anxiety symptoms in children with high-functioning autism spectrum disorders: A pilot study. *Focus on Autism and Other Developmental Disabilities*, *24*(1), 27–37. https://doi.org/10.1177/1088357608327666

Reaven, J., & Hepburn, S. (2003). Cognitive- behavioural treatment of obsessive- compulsive disorder in a child with asperger syndrome. *Autism*, *7*(2), 145–164.

Reeves BC, Deeks JJ, Higgins JPT, Wells GA. Chapter 13: Including non-randomized studies. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org

Reichow B., Volkmar, F. R. (2010). Social skills interventions for individuals with autism: evaluation for evidence-based practices within a best evidence synthesis framework. Journal of Autism and Developmental Disorders; 40:149-166.

Riva, G. (2005). Virtual Reality in Psychotherapy: Review. *CyberPsychology & Behavior*, *8*(3), 220–230. https://doi.org/10.1089/cpb.2005.8.220

Rodgers, J., Glod, M., Connolly, B., & McConachie, H. (2012). The relationship between anxiety and repetitive behaviours in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*(11), 2404–2409. https://doi.org/10.1007/s10803-012-1531-y

Rothbaum, Barbara Oiasov, Ready, D., Shahar, F., Grasp, K., Pair, J., Hebert, P., … Baltzel, D. (1999). Virtual Reality Exposure Therapy for PTSD Vietnam Veterans : A Case Study. *Journal of Traumatic Stress*, *12*(2), 263–271.

Rothbaum, Barbara Olasov, Hodges, L., Watson, B. A., Kessler, G. D., & Opdyke, D. (1996). Virtual reality exposure therapy in the treatment of fear of flying: a case report. *Behaviour Research and Therapy*, *34*(5–6), 477–481. https://doi.org/10.1016/0005-7967(96)00007-1

Russell, E., & Sofronoff, K. (2005). Anxiety and social worries in children with Asperger syndrome. *Australian and New Zealand Journal of Medicine*, *39*, 633–638.

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666–676. https://doi.org/10.1093/ije/dym018

Sansosti, F. J., & Powell-Smith, K. a. (2008). Using Computer-Presented Social Stories and Video Models to Increase the Social Communication Skills of Children With High-Functioning Autism Spectrum Disorders. *Journal of Positive Behavior Interventions*, *10*(3), 162–178. https://doi.org/10.1177/1098300708316259

Serret, S., Hun, S., Iakimova, G., Lozada, J., Anastassova, M., Santos, A., … Askenazy, F. (2014). Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new Serious game : a pilot study, 1–17.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ (Online)*, *349*(January), 1–25. https://doi.org/10.1136/bmj.g7647

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(8), 921–929. https://doi.org/10.1097/CHI.0b013e318179964f

Smith, L. E., Maenner, M. J., & Seltzer, M. M. (2012). Developmental trajectories in adolescents and adults with autism: The case of daily living skills. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(6), 622–631. https://doi.org/10.1016/j.jaac.2012.03.001

Smith, M. J., Fleming, M. F., Wright, M. A., Losh, M., Boteler, L., Dale, H., & Morris, O. (2015).

Brief Report: Vocational Outcomes for Young Adults with Autism Spectrum Disorders at Six Months After Virtual Reality Job Interview Training. *Journal of Autism and Developmental Disorders*, *45*(10), 3364–3369. https://doi.org/10.1007/s10803-015-2470-1

Smith, M. J., Ginger, E. J., Wright, K., Wright, M. A., Lounds, J., Laura, T., … Fleming, M. F. (2014). Virtual Reality Job Interview Training in Adults with Autism Spectrum Disorder, 2450–2463. https://doi.org/10.1007/s10803-014-2113-y

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Online)*, *355*, 4–10. https://doi.org/10.1136/bmj.i4919

Stichter, J. P., Laffey, J., Galyen, K., & Herzog, M. (2014). iSocial : Delivering the Social Competence Intervention for Adolescents (SCI-A) in a 3D Virtual Learning Environment for Youth with High Functioning Autism. *Journal of Autism and Developmental Disorders*, 417–430. https://doi.org/10.1007/s10803-013-1881-0

Strickland, D. (1998). Virtual Reality for the Treatment of Autism. *Virtual Reality in Neuro-Psycho-Physiology*.

Strickland, D. C., Coles, C. D., & Southern, L. B. (2013). JobTIPS: A transition to employment program for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(10), 2472–2483. https://doi.org/10.1007/s10803-013-1800-4

Strickland, D. C., Mcallister, D., Coles, C. D., & Osborne, S. (2007). An Evolution of Virtual Reality Training Designs for Children With Autism and Fetal Alcohol Spectrum Disorders, *27607*.

Sukhodolsky, D. G., Scahill, L., Gadow, K. D., Arnold, L. E., Aman, M. G., McDougle, C. J., … Vitiello, B. (2008). Parent-rated anxiety symptoms in children with pervasive developmental disorders: Frequency and association with core autism symptoms and cognitive functioning. *Journal of Abnormal Child Psychology*, *36*(1), 117–128. https://doi.org/10.1007/s10802-007-9165-9

Sze, K. M., & Wood, J. J. (2007). Cognitive behavioral treatment of comorbid anxiety disorders and social difficulties in children with high-functioning autism: A case report. *Journal of Contemporary Psychotherapy*, *37*(3), 133–143. https://doi.org/10.1007/s10879-007-9048-y

Sze, K. M., & Wood, J. J. (2008). Enhancing CBT for the treatment of autism spectrum disorders and concurrent anxiety. *Behavioural and Cognitive Psychotherapy*, *36*(4), 403–409.

https://doi.org/10.1017/S1352465808004384

Topol, E. (2019). *The Topol Review – Preparing the healthcare workforce to deliver the digital future*. Retrieved from https://topol.hee.nhs.uk/

Turner-Brown, L. M., Perry, T. D., Dichter, G. S., Bodfish, J. W., & Penn, D. L. (2008). Brief Report: Feasibility of Social Cognition and Interaction Training for Adults with High Functioning Autism. *Journal of Autism and Developmental Disorders*, *38*(9), 1777–1784. https://doi.org/10.1038/jid.2014.371

van Steensel, F. J. A., Bögels, S. M., & Perrin, S. (2011). Anxiety Disorders in Children and Adolescents with Autistic Spectrum Disorders: A Meta-Analysis. *Clinical Child and Family Psychology Review*, *14*(3), 302–317. https://doi.org/10.1007/s10567-011-0097-0

Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, Santaguida PL, Shamliyan T, Singh K, Tsertsvadze A, T. J. (2012). Assessing the Risk of Bias in Systematic Reviews of Health Care Interventions. *Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews*.

Walters, S., Loades, M., & Russell, A. (2016). A Systematic Review of Effective Modifications to Cognitive Behavioural Therapy for Young People with Autism Spectrum Disorders. *Review Journal of Autism and Developmental Disorders*, 137–153. https://doi.org/10.1007/s40489-016-0072-2

Wang, P., & Spillan, A. (2009) Evidence-based social skills interventions for children with autism: A meta-analysis. Ed Training Devel Disabil. 44, 318-342.

Wang, M., & Reid, D. (2011). Virtual Reality in Pediatric Neurorehabilitation: Attention Deficit Hyperactivity Disorder, Autism and Cerebral Palsy. *Neuroepidemiology*, *36*, 2–18. https://doi.org/10.1159/000320847

Waterworth, J.A., & Waterworth, E.L. (2001) 'Focus, Locus and Sensus: The 3 Dimensions ofVirtual Experience', Cyberpsychology & Behavior 4: 203–214.

Wells, G. A. S. B. (2001). The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. http://www. ohri.ca/programs/clinical_epidemiology/oxford. asp.

Weston, L., Hodgekins, J., & Langdon, P. E. (2016). Effectiveness of cognitive behavioural therapy with people who have autistic spectrum disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *49*, 41–54. https://doi.org/10.1016/j.cpr.2016.08.001

White, S. W. (2012). Growing pains: How psychologists can help to meet the clinical needs of clients with autism spectrum disorders. Cognitive and Behavioral Practice, 19(3), 433-436.

White, S. W., Ollendick, T. H., & Bray, B. C. (2011). College students on the autism spectrum: Preva- lence and associated problems. Autism: The Inter- national Journal of Research and Practice, 15, 683– 701. doi: 10.1177/1362361310393363

White, S., Oswald, D., Ollendick, T., & Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review*, *29*(3), 216–229. https://doi.org/10.1016/j.cpr.2009.01.003.Anxiety

White, S. W., Mazefsky, C. A., Dichter, G. S., Chiu, P. H., Richey, John, A., & Ollendick, T. H. (2014). Social-cognitive, physiological, and neural mechanisms underlying emotion regulation impairments: Understanding anxiety in autism spectrum disorder. *International Journal of Developmental Neuroscience*, *39*, 22–36. https://doi.org/10.1016/j.ijdevneu.2014.05.012.Social-cognitive

White, S. W., Richey, J. A., Gracanin, D., Coffman, M., Laconte, S., & Ollendick, T. H. (2016). Psychosocial and Computer-Assisted Intervention for College Students with Autism Spectrum Disorder: Preliminary Support for Feasibility. *Education and Training in Autism and Developmental Disabilities*, *51*(3), 307–317.

Williams White, S., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders*, *37*(10), 1858–1868. https://doi.org/10.1007/s10803-006-0320-x

Wing L. & Gould J. (1979) Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. Journal of Autism and Developmental Disorders 9, 11–29

Witmer BG, Singer MJ: Measuring presence in virtual environments: a presence ques- tionnaire. Presence 1998; 7: 225–240.

Wood, J. J., Drahota, A., Sze, K., Har, K., Chiu, A., & Langer, D. A. (2009). Cognitive behavioral therapy for anxiety in children with autism spectrum disorders: a randomized, controlled trial. *Journal of Child Psychology & Psychiatry*, *50*(3), 224–234. https://doi.org/10.1111/j.1469-7610.2008.01948.x.Cognitive

Wood, J. J., Ehrenreich-May, J., Alessandri, M., Fujii, C., Renno, P., Laugeson, E., … Storch, E. A. (2015). Cognitive Behavioral Therapy for Early Adolescents With Autism Spectrum Disorders and Clinical Anxiety: A Randomized, Controlled Trial. *Behavior Therapy*, *46*(1), 7–19. https://doi.org/10.1016/j.beth.2014.01.002

Wood, J. J., Fujii, C., & Renno, P. (2011). Cognitive behavioral therapy in high-functioning autism: review and recommendations for treatment development. In Evidence-based

practices and treatments for children with autism (pp. 197-230). Springer, Boston, MA.

Yang, Y. J. D., Allen, T., Abdullahi, S. M., Pelphrey, K. A., Volkmar, F. R., & Chapman, S. B. (2017). Brain responses to biological motion predict treatment outcome in young adults with autism receiving Virtual Reality Social Cognition Training: Preliminary findings. *Behaviour Research and Therapy*, *93*, 55–66. https://doi.org/10.1016/j.brat.2017.03.014

Yang, Y. J. D., Allen, T., Abdullahi, S. M., Pelphrey, K. A., Volkmar, F. R., & Chapman, S. B. (2018). Neural Mechanisms Of Behavioral Change In Young Adults With High-Functioning Autism Receiving Virtual Reality Social Cognition Training : A Pilot Study, (March), 713–725. https://doi.org/10.1002/aur.1941

Zeng, X., Zhang, Y., Kwong, J. S. W., Zhang, C., Li, S., Sun, F., … Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: A systematic review. *Journal of Evidence-Based Medicine*, *8*(1), 2–10. https://doi.org/10.1111/jebm.12141

APPENDIX A - DSM-V Diagnostic Criteria for Autism Spectrum Disorder

Must meet criteria A, B, C, and D:

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.

2. Deficits in nonverbal communicative behaviours used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.

3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behaviour to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

Specify current severity:

Severity is based on social communication impairments and restricted, repetitive patterns of behaviour (see Table 1).

B. Restricted, repetitive patterns of behaviour, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypes, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behaviour (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).

3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

Specify current severity:

Severity is based on social communication impairments and restricted, repetitive patterns of behaviour (see Table 1).

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.

Note: Individuals with a well-established DSM-IV diagnosis of autistic disorder, Asperger's disorder, or pervasive developmental disorder not otherwise specified should be given the diagnosis of autism spectrum disorder. Individuals who have marked deficits in social communication, but whose symptoms do not otherwise meet criteria for autism spectrum disorder, should be evaluated for social (pragmatic) communication disorder.

Specify if:

- With or without accompanying intellectual impairment
- With or without accompanying language impairment
- Associated with a known medical or genetic condition or environmental factor
- Associated with another neurodevelopmental, mental, or behavioural disorder
- With catatonia (refer to the criteria for catatonia associated with another mental disorder for definition)

APPENDIX B - Quality Assessment Questions

From Downs and Black, 1998

REPORTING

Questions answered: Yes (1) or No (0)

1. **Is the hypothesis/aim/objective of the study clearly described?**

   Are the main outcomes to be measured clearly described in the Introduction or Methods section? If the main outcomes are first mentioned in the Results section, the question should be answered no.

2. **Are the main outcomes to be measured clearly described in the Introduction or Methods section?**
   If the main outcomes are first mentioned in the Results section, the question should eb answered no.

3. **Are the characteristics of the patients included in the study clearly described?**

   In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.

4. **Are the interventions of interest clearly described?**

   Treatments and placebo (where relevant) that are to be compared should be clearly described.

5. **Are the distributions of principal confounders in each group of subjects to be compared clearly described?**

   A list of principal confounders is provided. Answer Yes (2), Partially (1), or no (0)

6. **Are the main findings of the study clearly described?**

   Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below).

7. **Does the study provide estimates of the random variability in the data for the main outcomes?**
   In non-normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.

8. **Have all important adverse events that may be a consequence of the intervention been reported?**

   This should be answered yes if the study demonstrates that there was a comprehensive attempt to measure adverse events. (A list of possible adverse events is provided).

9. **Have the characteristics of patients lost to follow-up been described?**

   This should be answered yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered no where a study does not report the number of patients lost to follow-up.

10. **Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?**

EXTERNAL VALIDITY

Questions answered: Yes (1), No (0) or Unable to determine (0)

All the following criteria attempt to address the representativeness of the findings of the study and whether they may be generalised to the population from which the study subjects were derived.

11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?

The study must identify the source population for patients and describe how the patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.

12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?

The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.

13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?

For the question to be answered yes the study should demonstrate that the intervention was representative of that in use in the source population. The question should be answered no if, for example, the intervention was undertaken in a specialist centre unrepresentative of the hospitals most of the source population would attend.

INTERNAL VALIDITY – BIAS

Questions answered: Yes (1), No (0) or Unable to determine (0)

14. Was an attempt made to blind study subjects to the intervention they have received?

For studies where the patients would have no way of knowing which intervention they received, this should be answered yes.

15. **Was an attempt made to blind those measuring the main outcomes of the intervention?**

16. **If any of the results of the study were based on "data dredging", was this made clear?**

    Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.

17. **In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?**

    Where follow-up was the same for all study patients the answer should yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.

18. **Were the statistical tests used to assess the main outcomes appropriate?**

    The statistical techniques used must be appropriate to the data. For example nonparametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.

19. **Was compliance with the intervention/s reliable?**

    Where there was non-compliance with the allocated treatment or where there was contamination of one group, the question should be answered no. For studies where the effect of any misclassification was likely to bias any association to the null, the question should be answered yes.

20. **Were the main outcome measures used accurate (valid and reliable)?**

    For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered as yes.

INTERNAL VALIDITY - CONFOUNDING (SELECTION BIAS)

Questions answered: Yes (1), No (0) or Unable to determine (0)

21. **Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?**

    For example, patients for all comparison groups should be selected from the same hospital. The question should be answered unable to determine for cohort and case control studies where there is no information concerning the source of patients included in the study.

22. **Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?**

    For a study which does not specify the time period over which patients were recruited, the question should be answered as unable to determine.

23. **Were study subjects randomised to intervention groups?**

    Studies which state that subjects were randomised should be answered yes except where method of randomisation would not ensure random allocation. For example alternate allocation would score no because it is predictable.

24. **Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?**

    All non-randomised studies should be answered no. If assignment was concealed from patients but not from staff, it should be answered no.

25. **Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?**

    This question should be answered no for trials if: the main conclusions of the study were based on analyses of treatment rather than intention to treat; the distribution of known confounders in the different treatment groups was not described; or the distribution of known confounders differed between the treatment groups but was not taken into account in the analyses. In nonrandomised studies if the effect of the main confounders was not investigated or confounding was demonstrated but no adjustment was made in the final analyses the question should be answered as no.

26. Were losses of patients to follow-up taken into account?

If the numbers of patients lost to follow-up are not reported, the question should be answered as unable to determine. If the proportion lost to follow-up was too small to affect the main findings, the question should be answered yes.

POWER

27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?

Sample sizes have been calculated to detect a difference of x% and y%.

Size of smallest intervention group Score

A <n1…. 0

B n1–n2…. 1

C n3–n4…. 2

D n5–n6…. 3

E n7–n8…. 4

F n8+…. 5

(Downs & Black, 1998)

APPENDIX C – Quality Assessment Scores

	Reporting										External Validity			Internal Validity-bias							Internal Validity-confounding						Power	App Total	Total	Score (%)
Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**			
Highest score	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**5**	**32**		
Paper																														
Amaral et al., 2018	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1	NA	NA	NA	NA	0	1	5	28	18	0.642
Austin et al., 2008	0	1	1	0	0	1	NA	0	1	NA	0	0	1	0	0	NA	1	NA	1	1	NA	NA	NA	NA	0	1	NA	25	8	0.32
Burke et al., 2018	1	1	0	1	2	1	1	0	1	1	0	1	0	0	1	1	1	1	1	1	NA	NA	NA	NA	0	1	5	28	22	0.785
Cheng et al., 2010	1	1	1	1	1	0	NA	0	1	NA	0	0	0	0	0	1	0	0	1	1	NA	NA	NA	NA	0	1	NA	25	11	0.44
Cheng & Huang, 2012	0	1	1	1	0	0	NA	0	1	NA	0	0	1	0	0	1	0	0	1	0	NA	NA	NA	NA	0	1	NA	25	8	0.32
Cheng et al., 2015	1	1	0	0	0	1	NA	0	1	NA	0	0	1	0	0	1	0	0	1	1	NA	NA	NA	NA	0	1	NA	25	9	0.36
Cheng et al., 2018	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	1	0	0	1	5	32	22	0.687
Cheng & Ye, 2010	1	1	1	0	0	1	NA	0	1	NA	0	0	1	0	0	1	0	0	1	1	NA	NA	NA	NA	0	1	NA	25	10	0.4
Didehbani et al., 2016	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	NA	NA	NA	NA	0	1	5	28	22	0.785
Herrera et al., 2008	1	1	1	1	0	1	NA	0	1	NA	0	0	1	0	0	1	1	1	1	1	NA	NA	NA	NA	0	1	NA	25	13	0.52

	Reporting										External Validity			Internal Validity-bias							Internal Validity-confounding						Power	App Total	Total	Score (%)
Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**			
Highest score	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**5**		**32**	
Paper																														
Ip et al., 2018	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0	1	5		32	20	0.625
Kandalaft et al., 2013	1	1	1	1	0	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	NA	NA	NA	NA	0	1	5		28	20	0.714
Ke & Im, 2013	1	1	1	1	1	1	NA	0	1	NA	0	0	1	0	0	1	1	0	1	1	NA	NA	NA	NA	0	1	2		26	15	0.576
Ke & Lee, 2016	1	1	0	1	0	1	NA	0	1	NA	0	0	1	0	1	NA	1	NA	0	1	NA	NA	NA	NA	0	1	1		26	11	0.423
Lorenzo et al., 2016	0	1	0	1	0	0	1	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	1	0	0	1	5		32	16	0.5
Lorenzo et al., 2013	1	0	0	1	0	0	1	0	1	NA	0	0	0	0	0	1	1	0	0	0	NA	NA	NA	NA	1	1	5		27	13	0.481
Maskey et al., 2014	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	NA	NA	NA	NA	0	1	5		28	22	0.785
Mitchell et al., 2007	1	1	1	1	2	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	3		32	21	0.656
Parsons et al., 2006	0	0	1	1	0	0	NA	0	1	NA	0	0	0	0	0	0	1	NA	1	0	NA	NA	NA	NA	1	1	1		26	8	0.307
Serret et al., 2014	1	1	1	1	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	NA	NA	NA	NA	0	1	5		28	20	0.714

	Reporting										External Validity			Internal Validity-bias							Internal Validity-confounding						Power	App Total	Total	Score (%)
Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**			
Highest score	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**5**	**32**		
Paper																														
Smith et al., 2014 and 2015	1	1	1	0	0	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	0	1	5	32	22	0.687
Stichter et al., 2014	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	NA	NA	NA	NA	1	1	5	28	21	0.75
Strickland et al., 2013	1	1	1	1	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	5	32	25	0.781
Yang et al., 2017	1	1	1	1	0	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	NA	NA	NA	NA	1	1	5	28	21	0.75
TOTAL	20	22	18	19	10	18	15	1	24	11	5	1	10	0	8	21	20	13	21	21	6	0	4	1	6			24	77	