Promoting identification of people with autism: Developing a behavioural screen for people with learning disability.

D Metcalfe

2020

Promoting identification of people with autism: Developing a behavioural screen for people with learning disability.

Dale Robert Metcalfe

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the Faculty of Health and Life Sciences, Psychology Department

September 2020

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by Northumbria University's Health and Life Sciences, ethics committee on 23rd November 2017.

I declare that the Word Count of this Thesis is 79,683 words

Name: Dale Robert Metcalfe

Signature:

Date: 26th September 2020

Written with sincerest thanks to the many organisations, the hundreds of participants and their families who all contributed volumes to this thesis.

Without all of you, this research would never have been possible.

Abstract

Autism is a lifelong developmental condition, diagnosed on the basis of persistent social communication deficits and repetitive, restrictive patterns of behaviour. Autistic people form a heterogenous group, but many experience challenges compared to non-autistic people. Identifying and diagnosing autistic people can be beneficial in helping them to better understand themselves, and to access support and interventions. Not all autistic people are diagnosed however, and one group which is disproportionately affected is people with learning disability. Screening tools have the potential to facilitate and speed up diagnosis by highlighting those who should be assessed in more detail. Many existing screening tools are largely inaccessible to people with learning disability due to the level of language required to complete them. The overall aim of the thesis is to facilitate the diagnosis and support of people with autism, in particular those with learning disability, by developing more accessible screening tools.

Chapter One provides an overview of the main issues to consider in the screening and diagnosis of autism, in those with and without learning disability. Chapter Two outlines the adaption of the Autism-Spectrum Quotient (AQ), with the aim of developing a more accessible version (AccAQ). The language of the AQ is simplified, and line drawings are added to enhance understanding. The results indicate that the original AQ and the adapted versions are equivalent in many ways, but the AccAQ still requires the person completing it to have some verbal communication, comprehension, and literacy skills. Chapter Three reports on a systematic review into the psychometric properties of autism screening tools used with people with learning disability. The review found that the majority of tools were lacking comprehensive validity and reliability data, were not specifically designed for use with people with learning disability, and required the input of an informant.

Chapters Four to Six discuss the development of a new autism screening tool, which was designed specifically for people with learning disability and does not require the input of a third-party informant. Chapter Four reports on the results of interviews with stakeholders about the current autism diagnostic pathway, the role of screening as a part of the diagnostic process and the properties that would be desirable in a new screening tool to ensure it is both accessible and useful. Overall, this study highlighted

the need for such a tool to have good psychometric properties, be clinically useful, user-friendly, cost-effective, and minimize demands on users and those being screened.

Chapter Five explored the literature that provided a theoretical basis for the inclusion of particular items within a behavioural screening tool. These are items that will prompt reactions which are indicative of the presence or absence of autistic traits: the concepts of empathy, mimicry, and contagion. These ideas are operationalised in a series of pilot studies which use existing video clips as stimuli and ask participants to self-report their responses to the videos. The results of a machine learning analysis, using cforest, showed that the self-reported responses were broadly predictive of the person's AQ score and provided proof of concept of the idea of a behavioural screening tool. Chapter Six extended this work. Here a set of custom stimuli were created and were viewed by autistic and non-autistic participants, with and without learning disability. Their reactions to these were self-reported and video recorded, and subsequently coded. The results indicated that, while scores on the behavioural screen could not discriminate between autistic and non-autistic participants, or predict self-reported AQ score, they did predict informant AQ scores.

Overall, the thesis makes a novel and significant contribution to the literature by identifying tools which are currently available to screen for autism in people with learning disability, alongside adapting the AQ to be more accessible to a wider range of people. It provides proof of concept for a behavioural screening tool for autism, which is specifically designed for people with learning disability, that integrates stakeholder views and theoretical literature during its initial development and testing.

Contents: Chapters, sections, and sub-sections.

Contents: Tables and figures.

0.0 Acknowledgements

First, I would like to thank my amazing supervision team. My principal supervisor Karen McKenzie, who has been an excellent mentor throughout this process. It has been a real privilege to have you share your knowledge with me, and to guide and counsel me when I encountered setbacks along the way. This support has made the thesis what it is today. Quite simply, I could not have asked for a better supervisor and look forward to working with you again in the future. I would also like to thank George Murray and the wider McKenzie-Murray family, who were always on hand to offer advice and guidance over the last three years. My tech whiz of a supervisor, Kris McCarty, whose skills and know-how allowed the tool to not only work, but to be something that participants genuinely enjoyed using. I cannot thank you enough for the support, and all the laughs along the way.

Second, I would like to give thanks to my much-appreciated friends and colleagues from Northumbria University Psychology Department, who have fostered an open and welcoming environment to work in. Specifically, Thomas Pollet who, for when all felt a little hopeless, found an amazing solution for my data. Not forgetting the best desk neighbour anyone could have wished for, Amy Newman. For all of the advice you have given me, the proof-reading you have offered to do for my thesis, and for ensuring that I stayed hydrated at all times. And the Lunch Club, who were not only great for bouncing ideas off, but who all made sure I took a break when I definitely needed to.

A massive thanks to my parents, Deborah Smith and Eric Metcalfe, who have supported me through every stage of my education – and who both helped to foster my curiosity from a very young age. And to my sister, Adele, who was happy to help out with my research in any way that she could.

Finally, I would like to thank Emily Carr. Since we met, you have always been very understanding of my late nights staring at spreadsheets, and all the weekends I have spent working on this thesis. And the 79,683 words it contains, that you stayed up late every night for an entire week to read…sorry about that. Thank you for actively engaging with my PhD – particularly throughout these final months, where your unconditional encouragement really kept me going.

Thank you.

1.0 Chapter 1: Introduction.

1.1 The background of autism

Autism, Asperger's syndrome, and autism spectrum disorder (ASD) are terms that refer to a lifelong condition which first exhibits itself in early childhood (American Psychiatric Association [APA], 2013). The history of the condition is complex, and references to it can be traced to folklore and fairy tales that originate from Germany and the UK. These stories draw on images of children who are replaced by changelings whose characteristics largely simulate children with autism; particularly those who regress in early stages of childhood (Leask, Leask, & Silove, 2005). There are medical accounts from the 1840s in the USA, relating to people who may fit the criteria for autism (Donvan & Zucker, 2016). One notable example is *Billy,* who although deemed 'intellectually incapacitated' had perfect musical pitch and could reproduce over 200 musical tunes.

The concept of autism was crystalized in the early 1900s during Eugen Bleuler's exploration of Schizophrenia (Stotz-Ingenlath, 2000). Bleuler, when describing his patients, defined autism as the detachment from outer reality with absolute predominance on their own inner life (Bizzari, 2018). The preoccupation with oneself, and the proposed ambivalence towards the world around them, went on to later influence thinking around what we now understand to be autism. Modern concepts of autism trace back to the 1940s via two separate strands of research, which both borrowed the terms 'autism' from Bleuler's original work. The first was a paper titled 'Autistic Disturbances of Affective Contact' by Leo Kanner (1943) which described eleven children with what he termed 'infantile autism,' a syndrome distinct from other already explored conditions. The second strand of research was by Hans Asperger who wrote independently about children with the same set of features; first writing about 'autistic psychopaths' in 1938, and then moving onto a more comprehensive study of children with the condition in 1944 (Asperger, 1991; Czech, 2018). In 1981, Asperger's articles were translated to English (Wing, 1981), leading to the idea of 'Asperger's syndrome,' which, while being divergent from autism, was very much connected to it.

In terms of clinical usage, changes in the Diagnostic and Statistical Manual (DSM) illustrate the changes which the concept of autism has undergone. In 1980, the DSM-III (APA, 1980) listed specific criteria

required for someone to meet in order to be diagnosed with autism: a lack of interest in people, severe impairments in communication, and bizarre responses to the environment, developing before 30 months old. In 1987, the DSM-III was updated (APA, 1987) and the criteria for autism were altered, with the addition of Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) for those who did not meet the stringent criteria for autism, but who still required support. Moreover, the strict age cut-off was removed. This was the first move toward the concept of the autistic spectrum. In the next iteration of the DSM, the DSM-IV (APA, 1994), Asperger's disorder was added, which can be considered to be toward the mild end of the spectrum. Finally, in the current DSM, the DSM-V (APA, 2013), the distinct categories within autism were collapsed and instead it was conceptualised as an all-inclusive diagnosis, ranging from mild to severe and dubbed the 'autism spectrum disorder'. For the purpose of this thesis, due to the frequent rejection of the term *disorder* by the autistic community (Fletcher-Watson & Happé, 2019), the term 'autism' will be used to refer to 'autism spectrum disorder' and any related conditions. As it is also the preference of the autistic community, identity first language (e.g. 'autistic people') will also be used throughout the thesis, except when discussing children explicitly, when terms such as 'children with autism' will be used, as this is preferred by families (Kenny et al., 2016).

1.1 Diagnosis, outcomes, and the prevalence of autism

As it currently stands, autism is diagnosed on the basis of persistent deficits in social communication and interaction, and restricted, repetitive patterns of behaviours, interests or activities (APA, 2013). The National Institute for Health and Care Excellence (NICE; 2016) outlines how a diagnostic assessment should be conducted, in which a comprehensive diagnosis is carried out by professionals who are competent and trained in autism diagnosis. At the core of this assessment is a multidisciplinary team, which may include: clinical psychologists, psychiatrists, nurses, speech and language therapists, occupational therapists, and support staff. The diagnosis will be informed by interviews with the person being diagnosed and family members, as well as information sourced from other stakeholders (e.g. their teacher). In many cases formal assessment tools may also be used, such as the Autism Diagnostic Interview – Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) or the Autism Diagnostic Observation

Schedule – Generic (ADOS-G; Lord et al., 2000). When considering a diagnosis, other potential coexisting conditions should be considered too, as these may prove influential.

Due to the broad criteria employed to identify and diagnose autism, there is large variation between autistic people. Communication difficulties for instance, can range from those who are non-verbal to those who simply flout conversational rules (Fletcher-Watson & Happé, 2019). Potential areas of differences between autistic and non-autistic people are vast and go beyond those listed in the diagnostic criteria. Notable differences include: understanding emotions of others (Harms, Martin, & Wallace, 2010), understanding the body's internal state or alexithymia (Shah, Hall, Catmur, & Bird, 2016), empathising with others (Baron-Cohen & Wheelwright, 2004; Jones, Happé, Gilbert, Burnett, & Viding, 2010), making eye contact (Dalton et al., 2005), sleep difficulties (Malow et al., 2016), motor difficulties (Licari et al., 2020), and sensorimotor difficulties including atypical reactions to stimuli (Hannant, Tavassoli, & Cassidy, 2016). In terms of daily living, autistic people may experience difficulties with meal times (Shmaya, Eilat-Adar, Leitner, Reif, & Gabis, 2017), social isolation (Humphrey & Lewis, 2008), bullying, (Church, Alisanski, & Amanullah, 2000; Humphrey & Lewis, 2008), increased anxiety (Croen et al., 2015), and increased suicidality (Cassidy & Rodgers, 2017; McDonnell et al., 2019). Due to this array of issues, autistic people often report stress and concern around unemployment (Hurlbutt & Chalmers, 2002), experience low rates of employment (The National Autistic Society, 2016), and may feel detached from the rest of society (Hurlbutt & Chalmers, 2002).

Aside from the impact on autistic people themselves, these difficulties may lead to negative outcomes for others. Parents of children with autism frequently report concerns about their children not being involved with their peers (Woodgate, Ateah, & Secco, 2008) and it is well documented that these parents also experience higher levels of stress compared to parents of typically developing children (Lovell, Moss, & Wetherell, 2012; Rao & Beidel, 2009). Adding to this, siblings report feeling burdened by responsibility and taking on a caring role from an early age (Dillenburger, Keenan, Doherty, Byrne, & Gallagher, 2010; Ferraioli & Harris, 2009). The societal impact of autism is great, with an estimated annual spending in relation to the unemployment of autistic people being £32 billion in the UK (Buescher, Cidav, Knapp, & Mandell, 2014).

It is clear that autism can lead to diverse outcomes for people and their families, and at times this can be particularly negative. One way in which the negative impact of autism can be mitigated is via intervention. Interventions are typically aimed at addressing a specific difficulty such as building social skills (Haworth, Libertus, & Landa, 2018; Rivard et al., 2016), alleviating psychological difficulties (e.g. depression and anxiety; Koudys, Perry, Ho, & Charles, 2018), or to address language difficulties (Bradshaw, Koegel, & Koegel, 2017). While such interventions can have success in improving quality of life and leading to more positive outcomes, in order to access interventions a person's autism must first be recognised and subsequently diagnosed.

The prevalence of autism, although unlikely to be growing, is being increasingly recognised. Between 1960 and 1980, autism was only thought to affect two to five children per ten thousand, whereas in the early 2000s it was thought to affect around ninety-four per ten thousand (Duchan & Patel, 2012), and now it is much higher at around one in one hundred people (Allison, Auyeung, & Baron-Cohen, 2012; Brugha et al., 2012). Once recognised, autistic people can be supported in various ways, yet specific subgroups are still underdiagnosed. One such group comprises those with co-occurring learning disability.

1.2 Learning disability

Learning disability, also known as intellectual disability, is a neurodevelopmental disorder that has a childhood onset, with the person experiencing significant difficulties with adaptive and cognitive functioning (APA, 2013). People with learning disability form a heterogeneous group with levels of severity from mild to profound. Classification of severity was previously determined by IQ (mild = 50-55 to 70, moderate = 35-40 to 50-55, severe = 20-25 to 35-40, and profound = less than 20 to 25: APA, 2000). Today, severity is classified in terms of adaptive functioning in order to better inform the support needs of those with learning disability, with severity and requirement for support increasing as daily living skills decrease (APA, 2013; World Health Organisation [WHO], 2018). Real-world challenges that people with learning disability face are varied, but communication difficulties are common (Bradshaw, 2001; Ziviani, Lennox, Allison, Lyons, & Del Mar, 2004), as well as poorer social abilities, reduced self-determination (Nota, Ferrari, Soresi, & Wehmeyer, 2007), and unemployment (Timmons,

Hall, Bose, Wolfe, & Winsor, 2011), all of which can adversely affect quality of life. In addition, behaviours that challenge are relatively common and are experienced by around one in every five or six people. It is recommended that positive, functional approaches should be employed to address behaviours that challenge (Murphy, 2017), with a focus on preventing incidents and improving quality of life (NICE, 2015). There is also the significant issue of health inequalities between people with and without learning disabilities, whereby, due to a range of reasons such as diagnostic overshadowing, the additional physical and mental health conditions of the former may be missed (Javaid, Nakata, & Michael, 2019). This is a serious issue, for which one suggested solution is the use of a standardised assessment tool such as the Health Equality Framework (Rooney, Harris, & Collins, 2018).

For the vast majority the cause of their learning disability is unknown, with an estimated 55-60% of people having no identifiable cause (Mefford, Batshaw, & Hoffman, 2012; Topper, Ober, & Das, 2011). For those where causes are identifiable, one prominent reason is thought to be a person's genetic makeup, where mutations in RNA and DNA have occurred (Khan et al., 2012). Examples of this can be seen in Down syndrome, which is caused by a genetic mutation of Trisomy 23 (Patterson, 1987), and Fragile X syndrome, which has been linked to the FMR1 gene (Hoogeveen & Oostra, 1997). Many additional genes have also been indicated as causal for learning disability (Gilissen et al., 2014). Aside from genetics, other factors have been linked to an increased chance of learning disability including maternal health conditions, such as diabetes and asthma, particularly when not properly managed (O'Leary et al., 2013), as well as obesity (Mann, McDermott, Hardin, Pan, & Zhang, 2013), and depression and bipolar disorder (Mcdermott, Mann, & Hardin, 2015). Other pre-natal factors that are potentially significant include: exposure to soil pollutants such as lead and mercury (McDermott et al., 2014), maternal tobacco smoking (Braun, Daniels, Kalkbrenner, Zimmerman, & Nicholas, 2009), maternal alcohol consumption (Chokroborty-Hoque, Alberry, & Singh, 2014), and birth and labour complications such as pre-term birth (Langridge et al., 2013). Maternal age is thought to be influential, with younger mothers having an increased chance of bearing children with a mild/moderate disability, while older mothers have an increased chance of having a child with a severe learning disability (Fairthorne, Langridge, Bourke, & Leonar, 2013).

1.3 Diagnosing learning disability

Many of those with co-occurring conditions, such as Down syndrome, are diagnosed as having learning disability early in their life (NICE, 2018). For others, diagnosis would be ascertained via a neurodevelopmental assessment. An overview of what this assessment should comprise of is provided by The British Psychological Society (BPS; 2015). Key features of a good-quality, valid assessment include a measure of both intelligence and of adaptive functioning, along with an assessment of the person's developmental history. A measure of intelligence or intelligence quotient (IQ) should be assessed in multiple components rather than a singular score. The only tool currently deemed suitable for this purpose in the UK is the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008) which includes measures of verbal comprehension, perceptual reasoning, working memory and processing speed, as well as the full-scale IQ measure. If this tool is not accessible to the person being assessed however, other measures can be considered for use. The second component, a measure of adaptive behaviour, assesses the conceptual, social, and practical skills that are learned and performed by people in their everyday lives. These measures are less clearly defined but involve assessing a person's performance on tasks within these domains. Two tools available for this are the Vineland Adaptive Behaviour Scales – Second Edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005) and the Adaptive Behaviour Assessment System, Second Edition (ABAS-II; Harrison & Oakland, 2003). Once a diagnosis is put in place, people are commonly able to access support for the varying difficulties that they may face.

Many people with learning disability also have comorbid conditions, which are varied but include, an increased rate of psychiatric and mental disorders (Emerson, Einfeld, & Ellis, 2011) that can adversely affect a person's life. One of the most highly reported comorbid conditions is autism and this dual diagnosis can have negative outcomes for the individual. The quality of life for those with both conditions is thought to be worse in terms of interpersonal relationships, social inclusion and physical wellbeing (Arias et al., 2018). Adding to this, behaviours that challenge are also more severe when autism symptoms are more severe (Matson & Rivet, 2007). It is thought that children with both autism and learning disability, compared to autism alone, respond more poorly to interventions targeting

expressive language, play ability, and language ability (Ben-Itzchak & Zachor, 2007), indicating that more specialised and effective interventions are required for the former group. The exclusion of people with learning disability in autism research is likely to be a contributing factor (Russell et al., 2019) which has led to interventions and drug treatments being less likely to account for co-occurring learning disability.

1.4 The intersection of learning disability and autism

Autism is frequently underdiagnosed in people with learning disability, although prevalence rates of the two together are estimated to be increasing over time (Matson & Shoemaker, 2009). La Malfa, Lassi, Bertelli, Salvini, and Placidi (2004) conclude that 40% of those with learning disability have autism and 70% of those with autism have learning disability. Similarly, when summarising previous research, Buescher, Cidav, Knapp, and Mandell (2014) estimate that between 40% and 60% of people with autism also have learning disability. This high comorbidity is not reflected in clinical diagnoses. One population study found that, when a full assessment was carried out, only half of those with learning disability who met the criteria for autism had previously been diagnosed (Saemundsen et al., 2010). Tonnsen et al. (2016) have more recently conducted research in the USA, finding an increased prevalence rate of autism among people with learning disability compared to reported levels within the general population: observing two hundred and thirty-five cases of comorbid autism and learning disability per thousand, compared to the previously reported eleven per thousand. This research is in line with higher estimates of prevalence published by Bryson, Bradley, Thompson, and Wainwright (2008). Together, these findings indicate that autism in people with learning disability is underdiagnosed and requires attention.

A potential explanation for this underdiagnosis is that, due to the heterogeneity of those with learning disability, assessment and diagnosis of autism is particularly challenging. With such a large degree of overlap between the symptoms of autism and learning disability, they are difficult to parse out from one and other; as learning disability often overshadows other conditions, it is possible for autism to be missed by a clinician (O'Brien & Pearson, 2004). Another complicating factor is that because of poorer literacy skills (for a summary see: Poncelas & Murphy, 2007) individuals may lack the skills to self-report symptoms or complete measures that assist clinicians in their diagnosis.

1.5 The usefulness of screening tools

A potential solution to the issue of underdiagnosis is screening. Screening is a method that helps differentiate between people who are likely and unlikely to have a particular condition, with those who screen positive being recommended for full diagnostic assessment (Glascoe, 2005). Screening can be differentiated from assessment in that it is designed to indicate the likely presence of a particular condition, whereas the assessment is a comprehensive process to clarify the nature of the condition and inform the diagnosis and subsequent intervention (Glascoe, 2005; Public Health England, 2019). Though screening tools do not give a full diagnosis, they can be advantageous for numerous reasons. One benefit is that they can be used by a wider range of people compared to diagnostic tools. The Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) is one such screening tool for autism. It does not require the person administering it to have a particular professional background or training, whereas the Autism Diagnostic Interview Revised (ADI-R), a diagnostic instrument, can only be used by an appropriately qualified person (Lord et al., 1994).

Screening measures are typically quicker to use than diagnostic assessment tools, making them more time efficient for both clinicians and the person being screened. As a result of both of these factors, there is recognition amongst some professional bodies (e.g. BPS, 2003) that there can be pragmatic reasons to use screening tools, namely to facilitate timely diagnosis. The importance of screening tools for autism has been argued by Allison et al. (2012) as they can help to ensure that health professionals can refer individuals presenting with autistic traits and behaviours for a full assessment from relevant specialists. In turn, this leads to more people being accurately diagnosed with autism and receiving the support they need. Aside from diagnosis, screening can be useful for research purposes as there are not always resources available to conduct a full diagnostic assessment of every participant. Here, screening tools can be used to indicate if a participant is likely to have a particular condition (McKenzie & Murray, 2015). The use of screening in this context could help address the issue of under-representation of people with learning disability and co-occurring autism in research.

Screening for autism in people with learning disability is however, a challenge. Presently, the only autism screening tool recommended by NICE (2012) is the AQ (Baron-Cohen et al., 2001). This is a self-

report tool, where respondents read each item and in turn indicate to what extent they agree with it. It is unsuitable for many people with learning disability who have few to no literacy skills, demonstrating the need for an appropriate screening tool that is accessible to people with learning disability.

1.6 Project goals

This PhD thesis aims to improve screening for autism, specifically in people with learning disability. This is likely to have a number of associated benefits, particularly for the individual in terms of subsequent diagnosis, support, and intervention. On a wider scale, this screening may also benefit autism research by increasing the possibility of including those with co-occurring learning disability. Specifically, this thesis intends to make the current NICE recommended screening tool, the AQ, more accessible to people with learning disability, and to understand what other screening options may be available for this group of people. A further aim is to develop a new screening tool that people with learning disability can access without requiring an informant to provide information. This thesis contains seven chapters and the specific aim of each chapter, along with the studies contained within, will be outlined in detail. This thesis begins with an adaption of the AQ as it is the only screening tool for autism recommended by NICE, it is widely used in practice, and has an extensive existing evidence base. The next chapter identifies and reviews potential alternative screening tools. Subsequent chapters outline the development, pilot testing, and refinement of a new behavioural screening tool.

It is worth noting here that some aspects of the thesis were affected by Covid-19. Primarily this impacted data collection for Chapter Six, as the researcher was unable to collect data in person from March 2020. This will be highlighted where applicable. The researcher feels that this limitation does not detract significantly from the important findings outlined here, and further data will be collected when it is possible to do so.

Electronic Supplementary Materials (ESM) have been uploaded to the Open Science Framework (OSF). If relevant information about a chapter is available on the OSF, a link to the ESM is denoted at the beginning of the chapter and within the chapter at specific points.

2.0 Chapter 2: Adapting the Autism-Spectrum Quotient.

ESM: https://osf.io/e3728/ Materials found in folder titled 'Chapter 2'

2.1 Introduction

2.1.1 The original Autism-Spectrum Quotient

The Autism-Spectrum Quotient (AQ) is a brief, self-report tool that measures the degree to which an adult of at least normal intelligence has characteristics associated with the autism spectrum, which Baron-Cohen, et al., (2001) deemed autistic-like-traits (hereafter, referred to as autistic traits). The AQ was developed in 2001 by Baron-Cohen and colleagues and at the time, no other brief, self-report instruments that detect autistic traits were available. The authors considered that such a tool may be useful for scientific reasons, such as assessing a person's autistic traits in a research scenario, as well as in applied contexts when used as a screening tool.

The items within the AQ were selected to measure traits outlined in the DSM-IV (APA, 1994) criteria for autism. Fifty items were included (e.g. 'I find making up stories easy') to reflect the respondent's preferences, rather than behaviour. According to Baron-Cohen et al. (2001), the AQ is structured as five subscales, with ten items in each, that correspond to the areas of functioning: Social Skills, Attention Switching, Attention to Detail, Communication, and Imagination. Responses are measured on a four point Likert scale (definitely agree, slightly agree, slightly disagree, definitely disagree) which are subsequently scored dichotomously as either indicating an autistic trait or not, with higher scores indicating the presence of a greater number of autistic traits.

Baron-Cohen and colleagues (2001) originally tested the AQ with four groups: adults with Asperger syndrome or high functioning autism, a non-autistic control group, a sample of Cambridge university students, and winners of the mathematics Olympiad. The authors found that 80% of the autistic participants, compared to only 4% of the non-autistic participants, scored thirty-two or higher. The parents of a subsample of participants also completed a forty-item version of the AQ as a third-party informant and were found to rate their children 2.8 points higher than their children rated themselves. The AQ was also shown to have good test-retest reliability. The authors concluded that the AQ was an

effective, brief, self-assessment tool for adults of normal intelligence, using a cut-point of thirty-two or higher to indicate people who are likely to be autistic.

2.1.2 **Alternative AQs**

Since then, the AQ has been used as both a research instrument and a screening tool, and several alternative versions have been developed to make it applicable to a wider range of people. In order to try to address the need for timely diagnosis (Koegel, Koegel, Ashbaugh, & Bradshaw, 2014) and the issues which many face in getting that diagnosis (Jones, Goddard, Hill, Henry, & Crane, 2014), both adolescent and child versions of the AQ have been created. Baron-Cohen, Hoekstra, Knickmeyer, & Wheelwright (2006) adapted the original AQ into a third-party informant measure that parents and carers could use in relation to adolescents. Research by Baron-Cohen et al. (2006) found the subscales of this version had satisfactory internal consistency, the lowest being .66. The original cut-off score of thirty-two was found to be suitable for this group, the youngest of whom was nine, but it was suggested that a lower cut-point of thirty might be advisable in future research. This research highlighted the possibility of using the AQ with a younger group, particularly as the AQ score appeared unaffected by age in this sample.

Subsequently, Auyeung, Baron-Cohen, Wheelwright, and Allison (2008) created a third-party informant version of the AQ which was targeted towards younger children. They retained the core of the original fifty items but adapted some to make them more applicable to the target group. The scoring of this version was on a four point Likert scale, where participants scored one to four on each item. Again, higher scores indicated more autistic traits and a cut-point of seventy-six (out of 150) was found to be optimal. Using the same structure as the adult version, the subscales were found to have satisfactory internal consistency, the lowest being .83, and the overall measure had good test-retest reliability. The authors noted that some items in the child version of the AQ required conversational skills, and therefore recommended it for use with people of average intelligence who had such abilities.

Short forms for each of the scales have also been developed. Based on data from four thousand participants, that included autistic adults, adolescents, and children, and comparable non-autistic participants, Allison et al. (2012) found ten items with the highest discrimination index between the groups. Using only these items, the area under the curve was found to be superior compared to the longer

version. The adult, adolescent, and child short-form AQs were all highly correlated with their respective longer versions, with the lowest reported correlation being .92 for the adults. For the shorter versions, the optimal cut-point was found to be six or more, indicating a high likelihood of the person being autistic. Allison and colleagues showed that the shorter versions are arguably superior in terms of sensitivity and specificity, compared to the longer versions. The one drawback however, is that this short version lacks subscales which may be useful to some.

2.1.3 **Subscales and factor analysis of the AQ**

While the original five AQ subscales were used in the adolescent (Baron-Cohen et al., 2006) and child adaptions (Auyeung et al., 2008), there is not a consensus on the validity of these subscales. Austin (2005) collected AQ responses from undergraduate students and volunteer adults, and found that a twenty-six item, three-factor structure (Social Skills; Details/Patterns; Communication/Mindreading) fit best. In a study of a Dutch translation of the AQ in the Netherlands, Hoekstra, Bartels, Cath, and Boomsma's (2008) found the best fit to be a two-factor structure, whereby the Attention to Detail subscale was retained, but the remaining four subscales (Social Skills; Attention Switching; Communication; Imagination) were combined into one general Social Interaction subscale. Stewart and Austin (2009) studied the AQ in a large student sample, although it was not investigated whether participants were autistic or not. A forty-three item, four-factor model was shown to be the best fit for the data collected (Socialness; Patterns/Attention to Detail; Understanding Others/Communication; Imagination).

Together, these studies illustrate the lack of a clear factor structure of the AQ. Kloosterman, Keefer, Kelley, Summerfeldt, and Parker (2011) investigated these divergent factor structures using AQ data from over five hundred participants, and found that the three and four-factor structures provided the best fit. However, no structures were found to be an adequate fit in this sample, leading to the development of yet another factor structure. Upon removal of poorly loaded items, a twenty-eight item, five subscale model (Social Skills [8]; Communication/Mind-Reading [5]; Restricted/Repetitive Behaviour [5]; Imagination [5]; Attention to Detail [5]) was found to fit best.

In subsequent research, further factor structures have been proposed. With a non-clinical sample of over one thousand people, Hurst, Mitchell, Kimbrel, Kwapil, and Nelson-Gray (2007) found a three-factor structure was superior (Social Skills; Details/Patterns; Communication/Mindreading) and provided improved internal consistency. Notably, this study did not include autistic people. By contrast, a study by Lau, Kelly, and Peterson (2013), which did include many autistic adults, found evidence for a thirty-nine item, five-factor structure (Sociability; Social Cognition; Interest in Patterns; Narrow Focus; Resistance to Change). As can be seen, these differ from those originally proposed by Baron-Cohen et al. (2001). A further factor analysis on the AQ was conducted by Murray, McKenzie, Kuenssberg, and Booth (2015) in both autistic and non-autistic adults. Evidence was found for a bi-factor model, where the majority of items were loaded onto a general factor and the remaining items were loaded onto an Attention to Detail subscale. This was similar to that proposed by Hoekstra and colleagues (2008) with the Dutch version of the AQ.

This body of research shows that the factor structure of the AQ is uncertain and may be influenced by the participant sample. Murray, Booth, McKenzie, and Kuenssberg (2015) found that when autistic and non-autistic groups complete the AQ, they are indeed indicating numbers of autistic traits. The authors advise caution however, when comparing autistic to non-autistic people based on their AQ scores as these groups respond in divergent ways.

2.1.4 **The AQ as a research instrument and a screening tool**

Despite the lack of consensus in relation to its factor structure, the AQ is used extensively in research with autistic people as a way to support the development of other relevant measures, such as 'Reading the Mind in the Eyes' Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) and the Empathy Quotient (Baron-Cohen & Wheelwright, 2004). As evidence indicates that autistic traits are distributed throughout the non-autistic population (Constantino & Todd, 2003), the AQ has also been used extensively with non-autistic populations. This includes research with families of autistic people, including parents (Wheelwright, Auyeung, Allison, & Baron-Cohen, 2010) and twin studies (Hoekstra, Bartels, Verweij, & Boomsma, 2007) as well as research investigating how autistic traits relate to other facets of personality, such as neuroticism, extraversion, and agreeableness (Austin, 2005).

A systematic review of non-clinical AQ research is provided by Ruzich et al. (2015), which focusses on the AQ score and its distribution within the non-autistic population. This work shows that the average score of a non-autistic individual is seventeen, that males on average score higher than females, and that while differences in AQ scores exist between studies, these appear to be random. As for the AQ's usefulness in the non-autistic populations, one study to note is by Murray, Booth, McKenzie, Kuenssberg, and O'Donnell (2014) which investigates measurement invariance on the AQ scores of autistic and non-autistic participants. Confirmatory factor analysis of the responses of each group revealed that, while the AQ indeed measures underpinning autistic traits, these traits are not continuously distributed throughout the population and instead form two distinct groups. This implies that while the AQ can be used with clinical and non-clinical populations for research purposes, caution is needed when directly comparing the AQ scores of the two groups.

The AQ has also been found to be a useful screening tool, and has been recommended for this purpose by NICE (2012). Woodbury-Smith, Robinson, Wheelwright, and Baron-Cohen (2005) conducted research in which one hundred people, who were due to be assessed for a diagnosis of autism, completed the AQ. They found that the AQ score was able to discriminate between those who would and would not go on to be diagnosed with Asperger's syndrome. Woodbury-Smith and colleagues recommended using a cut-point of twenty-six, as this was found to be more sensitive although less specific. Similarly, Booth et al. (2013) found the short-form AQ to be just as effective as the long-form at identifying autistic people.

2.1.5 **Rationale and purpose**

The current study aims to undertake a further adaption of the AQ, to create a screening tool for autism for people with learning disability that has robust psychometric properties. Although there is reportedly a high co-occurrence of autism and learning disability (La Malfa et al., 2004; Matson & Shoemaker, 2009) and numerous versions of the AQ exist, no version has been specifically developed for this group. The versions that do exist are not accessible to most people with learning disability due to the requirement for literacy skills. While this could be addressed by developing a third-party informant version of the AQ for young people with learning disability, who often have regular contact with their family (Lippold &

Burns, 2009), this would be more challenging for adults with learning disability, many of whom have infrequent contact with family members (Bigby, 2008). There has only been very limited research into the use of the AQ with people with learning disability. Kenny and Stansfield (2016) included some people with learning disability in their study, but this group was so small (autistic $N = 10$; non-autistic $N = 9$), that no meaningful insights abouts the usability of the AQ could be obtained. The current study will outline the adaption of the AQ to create an accessible version and analyse it in comparison to the original AQ.

The purpose of screening tools is primarily to detect people who are likely to benefit from a diagnosis. A good screening tool should be accessible, quick and easy to administer, and appropriate for the group it is used with (Glascoe, 2005). It should also have robust psychometric properties which include good reliability, internal consistency, test-retest, and interrater reliability where appropriate (Leung, Trevena, & Waters, 2012). A tool should also be a valid measure, showing concurrent validity with existing measures, sensitivity of 70-80% or higher and specificity of 80% or higher (Glascoe, 2005). If a tool does not meet and evidence these standards, it cannot be considered a robust instrument and is unlikely to be of use in a clinical environment.

The present research outlines the development of two versions of an accessible AQ, one which retains the four point Likert scale and the other which uses only dichotomous responses. The latter was trialled as the AQ is generally scored dichotomously, and by reducing the options and therefore the complexity, it may make the AQ more accessible for some participants. The two versions were assessed in terms of their psychometric properties, focussing on their reliability and validity, with comparisons made between the original AQ, third-party informant versions of the AQ, and whether or not a person is diagnosed as autistic. The original and adapted versions of the AQ also underwent exploratory factor analysis to add to the literature concerning the underlying factor structure of the AQ, and to yield an understanding of potential constructs that underpin the adapted version of the AQ. If this study proves successful, it would provide a screening tool that is accessible to more people with learning disability. This may aid the diagnostic process by speeding it up (McKenzie et al., 2015) and detecting those who otherwise may

have been missed (Dawson, 2016). It would also increase the opportunities to include more people with learning disability in AQ-based research.

2.2 Method

2.2.1 Design

The study employed a repeated measures design, where participants completed both the original AQ and an adapted 'Accessible AQ.'

2.2.2 Materials

2.2.2.1 Original AQ

Self-Report AQ

The original long-form, fifty-item AQ (Baron-Cohen et al., 2001) and the original short-form, ten-item AQ (Allison et al., 2012) were used in this study. For both, the original dichotomous scoring system was used, where participants respond on a four point Likert scale, but these responses are scored as one or zero. The long-form original AQ was scored in terms of the five subscales.

Third-party/Informant-rater AQ

In the original validation of the AQ (Baron-Cohen et al., 2001), one analysis included ratings provided by parents of the adults who took part. Similarly, other published versions developed for children and adolescents (Auyeung et al., 2008; Baron-Cohen et al., 2006) rely on parents or carers to complete them, on behalf of the person either being screened or taking part in research. In the current study, informant versions were used. While informant versions of the AQ for children and adolescents are already published, to the researcher's knowledge, a version for adults is not.

The version from Baron-Cohen et al.'s original article only included forty items, on the basis that the ten remaining were too subjective to be answered by a third party. The subsequent informant versions, however, do not omit these ten items. Given that there is a strong basis in the literature for fifty-item informant versions, a fifty-item, third-party variant answered about adults is used here. The wording of these items was changed from self-report to a third-party rating, 'I' was changed to, 's/he' and 'me' to,

'him/her.' An example of this is Item 7 which was originally worded, 'Other people frequently tell me that what I've said is impolite, even though I think it is polite' was changed to, 'Other people frequently tell her/him that what s/he has said is impolite, even though s/he thinks it is polite' on the informant version.

Adolescent and child AQs

For some participants, a third-party informant, who knows them well, completed either the long-form adolescent (Baron-Cohen et al., 2006) or child (Auyeung et al., 2008) versions of the AQ, or their shortened ten-item counterparts (Allison et al., 2012). The dichotomous scoring system was used.

2.2.2.2 Accessible adaptions of the AQ

Developing adaptions of the AQ

With permission from Simon Baron-Cohen, two versions of an Accessible AQ (AccAQ) were developed. As will be outlined in further detail below, Version A (AccAQ-A) asks participants to select a choice of two statements and Version B (AccAQ-B) retains the four point Likert scale.

The development of the AccAQ aimed to make the content of the AQ more accessible to people with mild to moderate learning disability. As such, it was informed by research and guidance on making information more accessible.

Townsley, Rodgers, and Folkes (2003) recommend that in order to simplify language, sentences should be kept short and only one idea should be present in each sentence; figurative language should be avoided; readers should be addressed directly; and straightforward everyday words should be used consistently, rather than synonyms. They also recommend that images, which aim to encapsulate the meaning of the words, are displayed to enhance understanding and ideally are placed beside the sentences they relate to.

Fajardo et al. (2014) found that higher sentence density leads to a reduced ability to find relationships between the sentences, resulting in poorer comprehension. They also argue that lengthy texts can look more complex and that this perception of difficulty may cause reduced motivation, leading to lower comprehension. Similarly, Sutherland and Isherwood (2016) show that having a clear format and less

complex text aids comprehension. These findings are echoed by Waight and Oldreive (2020) who demonstrate that using clear, straightforward, and jargon-free language can enhance accessibility.

In terms of the AccAQ, each item is numbered, and a box is placed around each item to break up the information, to help clarify where participants should respond. A clear format is used throughout to help ensure that the AccAQ is perceived as less difficult to read. Sentence density is largely out of the researcher's control as the AccAQ is an adaption of an existing scale. By breaking up the text and only requiring participants to read and comprehend one section at a time before answering a question, the aim is to reduce the density as much as possible. While the meaning of each AQ item is fixed, the language of the AQ can be simplified. Accordingly, the wording of each item has been made as clear as possible, for example, Item 18 was changed from, 'When I talk, it isn't always easy for others to get a word in edgeways' to, 'When I talk I forget to let other people have a turn.'

Townsley and colleagues recommend that supporting images are placed beside the relevant sentences. Unfortunately, this was difficult to do in the case of the AccAQ, and in this instance, the clearest way to present the images was by placing them below the sentences. This practice is common in other image-driven, accessible communication methods such as the Picture Exchange Communication System (PECS; National Autism Resources, 2020).

There has been a great deal of research on the inclusion of pictures in Easy Read resources, although the overall effectiveness of this appears to be mixed. The importance of using images, alongside other technologically backed resources (e.g. audio) has recently been highlighted, as these can enhance accessibility (Waight & Oldreive, 2020). When Poncelas and Murphy (2007) compared stand-alone simplified language to simplified language with pictures, they found that there was little benefit from adding pictures. The only exception being that repeated exposure to the pictures does confer an increased understanding of the content. Sutherland and Isherwood (2016) found that the inclusion of pictures leads to mixed results, though they also noted that familiarity with symbols is important as repeated exposure to an image can lead to increased comprehension.

However, findings to the contrary exist and should be acknowledged. The aforementioned studies by Poncelas and Murphy (2007), and Sutherland and Isherwood (2016) show that images are not always

helpful. Saletta, Kaldenberg, Rivera, and Wood (2019) also assessed the usefulness of different kinds of images, concluding that no particular type enhanced understanding and that using images does not promote content validity as adults do not frequently read picture books. Hurtado, Jones, and Burniston (2014) made the interesting finding that images alongside text lead to poorer comprehension than images alone, possibly due to the split of attention required by the audience. Chinn and Homeyard (2017) find that simply using pictures when presenting information appears to have little impact, but selecting images specific to individuals can have beneficial effects.

Many of these findings highlight that familiarity is an important construct (Chinn & Homeyard, 2017; Poncelas & Murphy, 2007; Sutherland & Isherwood, 2016), so where possible, images in the AccAQ have been used repeatedly. As an example the 'I am' image is always shown as a person pointing at themselves and can be seen in Figure 1. In the current study, it was not possible to make the materials specific to each person due to time and resource constraints. The decision to include images in the current study was taken as the overall findings are mixed, with some studies suggesting no real impact of their presence, while other researchers argue that they may be helpful. In addition, they can be used in isolation if the person chooses to ignore the wording, which may also be beneficial to confer greater understanding (Hurtado et al., 2014).

Callus and Cauchi (2020) argue that using Easy Read styled information does not guarantee people can actually access information. They demonstrate that the wider context within which this information is presented needs to be acknowledged, for example, whether there is someone available to structure and mediate the interaction. Indeed, Buell, Langdon, Pounds, and Bunning (2020) found that simplifying language alone does not enhance understanding. Instead, they found that improved comprehension resulted when the straightforward language of the Easy Read material was combined with a mediator to aid the person's interaction with it. Moreover, Hurtado et al. (2014) found that reading documents aloud to someone with learning disability can help them better understand their meaning.

This literature means that creating an accessible version of the AQ may not be enough for people with learning disability. In some circumstances, the person completing the tool may further benefit from having the items read to them, or having the ideas represented within the items explained to them in a

different way. For this reason, in the current study, if people wanted to have assistance completing the AccAQs or have items read to them by a friend, family member, support worker, or the researcher if they were present, this was encouraged.

The AccAQs were presented both online and on paper, and the items were identical in each. In line with Townsley et al.'s recommendations, presentation was optimised to ensure the text is legible and that the images are clear in both the online and paper versions. By employing past research to inform the adaption of the original AQ, it ensures that, in so far as possible, the original AQ content is retained and the AccAQs are accessible to people with learning disability. The first stage of the adaption of the AQ was the wording. In line with the recommendations outlined previously, the wording was simplified and made as accessible as possible. For example, Item 48: 'I am a good diplomat' was changed to, 'I am good at helping people get on with each other.' Wording changes were first made by the primary researcher, then edited by their supervisor, before being agreed upon by the whole research team.

The next stage was creating images to accompany each newly worded item. Images were commissioned from a graphic designer, who provided line drawings relevant to each item. The image for the person indicating 'I' is consistent and used in each item. Images have been added where appropriate and have been repeated where words are repeated; for example, each time the word 'good' is mentioned, a 'thumbs up' image is used. A copy of the original AQ, the wording of the adapted AccAQ, and ideas for images were given to the graphic designer who then provided draft images. These were discussed and edited by the research team and feedback was provided to the graphic designer. This refinement process continued until the final images used in this study were agreed. Examples of the items and their corresponding images can be seen in Figure 1.

Each version contains fifty items, presented in the same order as the long-form original AQ. A short-form version for each was also developed, using the accessible version of the items contained in the short-form adult version of the AQ. The reasons for this are discussed in the procedure section. The final stage was to present these on paper and online. As previously stated, the researcher ensured text and images were clearly visible in all cases.

As aforementioned, two versions of the AccAQ were created:

Accessible AQ Version A (AccAQ-A)

AccAQ-A builds on the dichotomous nature of the AQ by presenting the participant with two statements and accompanying images for each item. One statement indicates the presence of an autistic trait and the other does not. The participant was asked to read each statement and select the one they think is most applicable to them. Participants completing the paper copy were instructed to tick next to the most applicable statement, and those completing it online were asked to click on the most applicable option.

Accessible AQ Version B (AccAQ-B)

AccAQ-B is more similar to the original AQ. In this version, one statement is given, and the person responds on a four point Likert scale with accompanying images. The response of 'definitely agree' is depicted by two ticks, 'slightly agree' by one tick, 'slightly disagree' by one cross and 'definitely disagree' by two crosses. As with the original AQ, some items are reverse scored. However, it should be stated that others could not be reversed without impacting accessibility. As an example, Item 28 was originally worded, 'I usually concentrate more on the whole picture, rather than the small details' which is difficult to word in a more accessible way. As such, the final wording for this item was, 'When I look at a picture I look at the small details.' Complete versions of each AccAQ are available in the ESM (Chapter 2 / AccAQ Versions).

Example: AQ Item 26

Original AQ Wording: ‘I frequently find that I don’t know how to keep a conversation going.’

Accessible AQ Version A

Accessible AQ Version B

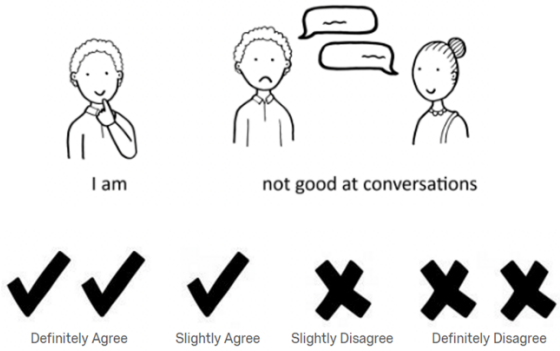


Figure 1: Example AccAQ items.

2.2.3 **Recruitment**

Participants were recruited via multiple routes. For the most part, participants were an opportunity sample recruited through word of mouth, and adverts posted on social media, online forums, and the university participant pool. The latter offers participants credits that can be used by them to reward others for participating in their own research. Other participants completed the AQ and the accessible versions as part of another study (see: Chapter Six for details). This recruitment was more targeted toward recruiting autistic people, people with learning disability, along with comparable controls. Recruitment for this aspect of the study was cut short due to the restrictions imposed as a result of Covid-19.

As part of the recruitment advert, participants were given some information about the original AQ, and were told that the study aimed to develop a version that was more accessible to people with learning disability. Participants were made aware that they would be asked to complete a series of questionnaires, either online or in person, depending upon their preference. They were provided with contact details for

the researcher and were directed to the online questionnaire page; if they preferred to complete the study on paper, they were directed to contact the researcher.

2.2.4 **Procedure**

This study received ethical approval from Northumbria Health and Life Sciences ethical approval committee through the online ethics portal.

Participants who were interested in the study were provided with full information about the research including background information, the purpose of the study, what they would be asked to do and what would happen to the data that they provided. Each participant was given the opportunity to ask questions, either via email, telephone, or if they preferred, in person. For participants under the age of 18, parents/carers provided consent on behalf of the participant. All participants aged 18 or over provided informed consent for themselves. Participants who took part in the study online gave informed consent via a tick-box, while those who completed it on paper provided written consent.

In the participant information, participants were made aware that someone else could read the questions to them. After consent was gained, participants' biographical information was provided by either the participant or a parent, in the case of children. This included age, gender, level of education, ethnicity, and any relevant diagnoses including autism and learning disability. Following this, participants completed the original AQ and were randomly assigned to complete AccAQ-A or AccAQ-B. Participants were then given the option of providing an email address to enable them to be invited to complete a follow-up questionnaire twenty-eight days later. Those who did give contact information were asked to complete both the AccAQ-A and AccAQ-B in a random order, and were asked which version they preferred along with the reasons behind their decision. All participants were given the opportunity to provide further comments.

In the more targeted recruitment process, which formed part of a larger study that included other procedures and questionnaires (see: Chapter Six), it quickly became apparent that asking some participants to complete the original AQ and one version of the AccAQ was too time-consuming. Due to this, the decision was made to develop and use the short-form versions of the original and AccAQ for

such participants. The AccAQs were presented and completed in a random order. During the targeted recruitment, the researcher completed the questionnaires with the participant. It was up to the participant how involved the researcher was, but often this included reading the items aloud to them or discussing items with the person so that they better understood the meaning behind each item.

Where possible, as part of this targeted recruitment strategy, third-party informants (e.g. parents, carers, partners, friends) were asked to complete a long-form AQ about the participant. These informants also provided some biographical details about themselves including their age, gender, and relationship to the person taking part.

Participants who completed the study online did so via the Qualtrics survey website and the remainder completed paper versions of the questionnaires. These collective results were imported or typed up into a database for subsequent analysis.

2.2.5 **Participant summary**

Participants were divided into those who completed the long-form original AQ and AccAQ, and those who completed the short-form versions of each. Those who completed the short-form versions usually completed both the AccAQ-A and AccAQ-B. Participants were required to have normal or corrected-to-normal vision so that they could see the items on the AccAQs. Participants were allowed to ask someone to help read the items if required. Exclusion criteria for the study were: anyone under five years old, anyone who is non-verbal, or anyone 18 or over without the capacity to consent for themselves. Consent for participants under 18 years old was provided by a parent or carer.

Long-form AQ (AQ-50)

In total, 477 people started the online questionnaire. Upon removal of those who did not finish and participants with five or more incomplete items, on any version of the AQ that they completed (criteria in line with Auyeung et al., 2008; Hoekstra et al., 2007), 411 participants remained. Twelve additional participants were included in this dataset from the more targeted sampling procedure, giving a total of 423 participants who completed the full-length version of the AQ.

The sample reported a mean age of 22.11 years old (*SD* = 7.70; minimum = 6, maximum = 67), 15 (3.6%) participants were under 18 years old. Seventy-eight (18.4%) participants identified as male, 339 (80.1%) as female, 5 (1.2%) as other (none specified further), and 1 (0.2%) did not report a gender. Of the 423 participants, 33 (7.8%) stated that they were autistic of which 4 (12.1%) reported having learning disability (1 mild, 3 moderate). Sixteen (48.5%) of these reported an additional diagnosis including learning difficulty, depression, anxiety, obsessive compulsive disorder (OCD), attention deficit hyperactive disorder (ADHD), post-traumatic stress disorder (PTSD), and dyspraxia. In the overall sample, 2 (0.5%) further people reported having learning disability (1 mild, 1 moderate) and 32 (7.6%) reported additional diagnoses including learning difficulty, depression, anxiety, OCD, ADHD, PTSD, dyspraxia, behavioural issues, and borderline personality disorder.

Overall, 227 (53.7%) people completed AccAQ-A, and 224 (53.0%) completed AccAQ-B. People were invited to complete the AccAQ again at a later time to assess test-retest reliability, and 117 (27.7%) participants did this, all of whom also completed both the AccAQ-A and AccAQ-B.

Short-form AQ (AQ-10)

Data for the short-form AQ was collected as part of a larger study with targeted recruitment, as reported in Chapter Six. A total of 158 participants (including 41 with learning disability) were included in the analysis in Chapter Six, however, not all participants completed the AccAQ. There were a number of reasons people did not, including not wanting to, or feeling they were unable to complete the AccAQ. In total, 88 participants completed the short version of the AccAQ through this study, which in turn means 70 did not complete the AccAQ during this study.

Under-18

Thirty-three (37.5%) participants were under 18 years old. The mean age of the group was 13.55 (*SD* = 2.10; minimum = 9; maximum = 17), 20 (60.6%) were male and 13 (39.4%) were female. Eight (24.2%) were neither diagnosed with autism nor suspected of being autistic and of these, 5 (15.2%) were diagnosed with learning disability (1 moderate, 4 severe). Two (6.1%) people reported additional diagnoses (ADHD and moderate hearing loss / epilepsy and chiari malformation type 1). Twenty-three

(69.7%) reported a diagnosis of autism and of these, 4 (17.4%) had a co-occurring learning disability. Fourteen (60.9%) were reported to have additional diagnoses including ADHD, dyspraxia, hypermobility, anxiety, and dyslexia. Two (6.1%) people reported that they suspected that they were autistic; either being on a waiting list for assessment or parents reporting that they had autistic traits. Neither of these reported a diagnosis of learning disability, 1 (50.0%) reported a diagnosis of dyspraxia. Twenty-seven (81.8%) of the participants also had information provided by a third-party informant, all of whom were parents.

All, except 2 (6.1%), of these participants completed the short self-report AQ, and everyone completed both AccAQ-A and AccAQ-B.

Over-18

Fifty-five (62.5%) participants were 18 years old or more. The mean age was 37.58 years (*SD* = 13.90; minimum = 18; maximum = 81), 22 (40.0%) participants were male, 31 (56.4%) female, and 2 (3.6%) other (not specified).

Twenty-one (38.2%) were neither diagnosed with autism nor suspected of being autistic. From these, 9 (42.9%) reported having learning disability (5 mild, 1 moderate, 3 unreported) and 7 (33.3%) reported an additional diagnosis including Down syndrome, anxiety, depression, ADHD, dyslexia, and bipolar disorder. Twenty-five (45.5%) reported being diagnosed with autism. Of these, 4 (16.0%) reported a co-occurring diagnosis of learning disability (2 mild, 1 severe, 1 unreported) and 16 (64.0%) reported an additional diagnosis including anxiety, ADHD, dyslexia, dyspraxia and epilepsy.

Nine (16.4%) participants stated that they had not been diagnosed but were suspected of being autistic and were awaiting assessment. None were diagnosed with learning disability, but 5 (55.6%) reported an additional diagnosis including anxiety, depression, PTSD, and borderline personality disorder.

All completed the original AQ, 53 completed the AccAQ-A and 53 AccAQ-B. Twenty-two (40.0%) of the participants over 18 years old also had information provided by an informant (8 parents, 13 children, 1 sibling).

2.2.6 **Analysis plan**

Data from the AQ was scored using the original dichotomous scoring method (Baron-Cohen et al., 2001), where higher scores indicate more autistic traits (out of a possible range of 0 to 50 for the long-form and 0 to 10 for the short-form). For data with fewer than five missing items, the AQ score was adjusted using the following calculation:

Total AQ score + (Mean item score x Number of missing items) = Corrected AQ score.

This was done for both the original and accessible versions. This method has been used in previous AQ research (Auyeung et al., 2008; Hoekstra et al., 2007). Missing datapoints for specific items were left blank, only total scores were corrected.

Using these scored long-form and short-form datasets, participants were split into age 18 or over (18+) and under 18s. These two groups were analysed separately as the original self-report variant of the AQ has only been validated for use in adult samples. Multiple analyses were run on the long-form and short-form AQs, on different sub-groups, as outlined in Table 1.

Additional analyses were run, these are detailed below for the long-form and short-form versions separately.

Long-form AQ (AQ-50)

For the 18+ group, Exploratory Factor Analysis (EFA) was performed on all long-form AQs. This includes the original AQ, as no consistent factor structure has been found in previous research. For the same reason, confirmatory factor analysis was not run. EFA began with preliminary checking and visualisation of the data, using both scatterplots and correlations between all items on each version of the AQ. The focus of this study was not to reduce the number of items, as this has already been done via the short-form version. However, items which indicated redundancy and multicollinearity, by having correlations of .9 or greater, were removed to prevent perfect multicollinearity. To identify the optimal number of factors present in the AQ, Parallel Analysis with Principal Components Analysis (PA-PCA) was employed. In PA-PCA, random eigenvalues are generated for the purpose of comparison to the eigenvalues produced from PCA of the data in question. The optimal amount of factors is identified as

the number found in the data, which have eigenvalues greater than the percentile of the corresponding randomly generated eigenvalues (O'Connor, 2000).

When conducting EFA, fit was forced to the optimal numbers of factors revealed in PA-PCA. As previous research shows that factors and subscales on the AQ are frequently correlated, Promax rotation was employed in this analysis to allow factors to remain correlated (Hurst et al., 2007).

Short-form AQ (AQ-10)

Subsequent to the analyses outlined in Table 1, a further set of receiver operating characteristic (ROC) curve analyses were run. In these analyses, short-form scores were derived from the long-form responses and combined with the short-form responses. ROC analyses were run on this larger dataset. The most appropriate cut-point will be identified along with the sensitivity and specificity.

Throughout the results section, different sample sizes are reported which reflects that some participants did not complete every element of the study. Using the results of the ROC curve analysis, a power analysis was run to ascertain whether the sample size was sufficient for this study. Additionally, a power analysis for Cronbach's alpha was run for the ten-item versions of the AccAQ.

Learning disability only analyses

The original plan was to conduct a separate set of analyses for people with learning disability. However, with the current sample, it was not possible to do so in a reliable manner. In total, only 28 participants reported having learning disability across the long-form ($N = 6$) and short-form ($N = 22$) data collection, reduced to 19 when the under-18 group is removed. This is insufficient data to reliably analyse the effectiveness of the AccAQ in people with learning disability.

All analyses were run in SPSS Statistics Version 24. Relevant syntax, outputs, and datafiles are available in the ESM (Chapter 2 / Analysis).

Table 1: Analyses on the long-form and short-form AQs, broken down by group.

	Long-form AQ (AQ-50)		Short-form AQ (AQ-10)	
Statistical Approach	18+	Under-18	18+	Under-18
Descriptive statistics	Including skewness and kurtosis	No skewness and kurtosis due to small sample	Including skewness and kurtosis	
Concurrent validity Correlation	AQ vs AccAQ AccAQ-A vs. AccAQ-B Original subscale on AQ vs subscales on AccAQ	AQ vs AccAQ	AQ vs AccAQ AccAQ-A vs. AccAQ-B Informant (long and short-form) AQ vs AQ and AccAQ	
Test-retest Reliability Correlation	AccAQ vs (same) AccAQ 28 days later*	-	-	-
Inter-item agreeability Kappa (interpreted according to Landis and Koch (1977)**	Specific items on AQ vs AccAQ	-	Specific items on AQ vs AccAQ	
Internal consistency Cronbach's alpha	Calculated for AQ and AccAQ	-	Calculated for AQ and AccAQ	
Preference Chi Square	Inspected for significance to ascertain preference for a specific AccAQ	-	Inspected for significance to ascertain preference for a specific AccAQ	
Screen accuracy Receiver operating characteristic (ROC) curve	-	-	Cut-point Sensitivity Specificity	

Note. Where only AccAQ is used, this refers to both AccAQ-A and AccAQ-B; where '-' is shown, this indicates that the analysis was not run.

*Note. Participants were invited, via email, to take part 28 days later. Some did not complete it on that day.

**Note. Item specific calculations are available in the ESM (Chapter 2 / Appendix Chapter 2).

2.3 Results

2.3.1 Long-form AQ

2.3.1.1 Concurrent validity

The 18+ group was separated from the under-18 group for the purpose of these analyses. Descriptive statistics, skewness, and kurtosis were calculated for the long-form original AQ, AccAQ-A, and AccAQ-B (see Table 2). Skewness was found to be positive for all versions, which indicated a greater number of participants scoring at the lower end of the scales. Kurtosis showed a normal distribution.

Table 2: Descriptive statistics for the long-form AQ, AccAQ-A, and AccAQ-B.

	Mean *(SD)*	Min – Max	Skewness *(SE)*	Kurtosis *(SE)*
Original AQ $N = 408$	18.67 (8.85)	3 – 46	1.00 (.12)	.53 (.24)
AccAQ-A $N = 217$	17.88 (10.08)	0 – 47	1.05 (.56)	.562 (.33)
AccAQ-B $N = 214$	16.91 (9.17)	0 – 46	1.146 (.17)	.96 (.31)

Scores of participants on the original AQ were plotted against their respective scores on AccAQ-A and AccAQ-B independently. Participants who completed the follow-up section of the study also had their scores on AccAQ-A plotted against the AccAQ-B. These plots outlined a strong linear relationship between the original and each AccAQ version indicating that, although the variables may be somewhat skewed, using a Pearson correlation would be a sensible approach. To be certain that these results were not influenced by the aforementioned skew, correlations were repeated using both a Spearman's Rho (r_s) correlation and a Pearson correlation on Log10 transformed data.

The Pearson correlations showed a strong significant relationship between the original AQ and both AccAQs, as well as a strong relationship between both AccAQs. These results can be found in Table 3.

Table 3: Correlations between the long-form original AQ, AccAQ-A, and AccAQ-B.

	AccAQ-A	AccAQ-B
Original AQ	$r = .95$ $p < .001$ $r_s = .92$, $p < .001$ $r_{(lg10)} = .91$, $p < .001$ $N = 217$	$r = .94$, $p < .001$ $r_s = .89$, $p < .001$ $r_{(lg10)} = .88$, $p < .001$ $N = 214$
AccAQ-A	-	$r = .98.$, $p < .001$ $r_s = .97.$, $p < .001$ $r_{(lg10)} = .95$, $p < .001$ $N = 114$

Participant scores on each of the original 5 subscales proposed by Baron-Cohen et al. (2001) were calculated for the original AQ and for the corresponding items on both AccAQs. The descriptive statistics for each subscale can be found in Table 4. Plotting participant scores on each subscale of the original AQ against the AccAQ version indicated linear relationships for all. Pearson correlations were then conducted to ascertain if each pair of subscales were related. These found strong relationships between the original AQ subscales and the corresponding subscales on each version of AccAQ. Results of these correlations can be found in Table 5.

Table 4: Descriptive statistics for each subscale of the long-form original AQ, AccAQ-A, and AccAQ-B.

	Mean *(SD)*				
	Social Skills	Attention Switching	Attention to Detail	Communication	Imagination
Original AQ $N = 408$	3.26 (2.45)	4.74 (2.56)	4.71 (2.33)	2.76 (2.52)	3.19 (1.80)
AccAQ-A $N = 217$	2.77 (3.15)	5.06 (2.74)	5.31 (2.43)	2.10 (2.64)	2.63 (2.38)
AccAQ-B $N = 214$	2.50 (2.80)	4.96 (2.36)	4.97 (2.40)	2.16 (2.39)	2.30 (2.07)

Table 5: Correlations between subscales on the long-form original AQ with corresponding subscales on the AccAQ-A and AccAQ-B.

	Original AQ				
	Social Skills	Attention Switching	Attention to Detail	Communication	Imagination
AccAQ-A $N = 217$	$r = .88, p < .001$	$r = .82, p < .001$	$r = .83, p < .001$	$r = .89, p < .001$	$r = .81, p < .001$
AccAQ-B $N = 213$	$r = .88, p < .001$	$r = .81, p < .001$	$r = .84, p < .001$	$r = .88, p < .001$	$r = .80, p < .001$

2.3.1.2 **Test-retest reliability**

For those participants who opted to complete the AccAQ on a second occasion, test-retest reliability was calculated. Participant scores from the original AccAQ and the respective version at retest were plotted against each other which showed a linear relationship. Pearson correlations were then run. Due to the aforementioned skew, Spearman's Rho and Pearson correlations on Log10 transformed data were also conducted. Pearson correlations were found to be good for both the AccAQ-A and AccAQ-B, which indicates that both have good test-retest properties. Spearman's Rho and the Pearson correlations on the Log10 transformed data further confirmed this. Results of the test-retest correlations can be found in Table 6.

Table 6: Correlations to investigate the test-retest ability of the long-form AccAQ-A and AccAQ-B.

	AccAQ-A	AccAQ-B
Follow-up Mean *(SD)*	18.55 (11.34)	18.55 (10.29)
Correlation	$r = .93, p < .001$ $r_s = .93, p < .001$ $r_{(lg10)} = .92, p < .001$ $N = 57$	$r = .93, p < .001$ $r_s = .93, p < .001$ $r_{(lg10)} = .91, p < .001$ $N = 64$

2.3.1.3 **Inter-item agreement**

To assess how well each item on the original AQ was related to the corresponding items on the AccAQs, Kappa was calculated between each set of items and the results were interpreted according to Landis and Koch (1977). The results showed that all items were significantly related to one another. Moderate agreement, or better, was shown for 41 items on the AccAQ-A and 42 items on the AccAQ-B. A summary of levels of agreement can be found in Table 7 and raw Kappa values for each item are presented in the ESM (Chapter 2 / Appendix Chapter 2).

Table 7: Kappa agreement between items on the long-form original AQ and each AccAQ variant.

Level of agreement (Landis & Koch, 1977)	AccAQ-A	AccAQ-B
No agreement (< 0)	2	2
Slight (0 — .20)	0	1
Fair (.21 — .40)	7	5
Moderate (.41 — .60)	21	16
Substantial (.61 — .80)	19	20
Perfect (.81–1.0)	1	6

2.3.1.4 **Internal consistency**

For each 50-item version of the AQ, Cronbach's alpha was calculated to investigate internal consistency. All versions were shown to have Cronbach's alpha values that fell within the 'good' range (Cohen, 2008). The results are shown in Table 8.

Table 8: Cronbach's alpha results for each long-form AQ variant.

	Original AQ	AccAQ-A	AccAQ-B
Cronbach's alpha	.89	.92	.91

2.3.1.5 **Preference**

Participants who completed the follow up component of the study, and who completed both the AccAQ-A and AccAQ-B, were asked which accessible version of the AQ they preferred. The results are shown in Table 9. Removing participants who had no preference, this data was analysed using a Chi-Square test. This indicated that no version of the AccAQ was significantly preferred over the other; $X^2(1, N = 98) = .37$, $p = .54$.

Table 9: Long-form AccAQ preferences.

	AccAQ-A	AccAQ-B	Neither
Number who prefer	46	52	15

2.3.1.6 **Under-18 sub-section**

As only 15 participants under 18 years old completed the long-form version of the original AQ and AccAQ, the sample is notably underpowered, and consequently only concurrent validity was assessed. Further analyses relevant to people under 18 will be carried out using the short versions of the AQ.

Participant scores were visualised on a scatterplot which indicated a linear relationship. Pearson correlations were then carried out, which indicated good agreement between the original version of the AQ and both adapted accessible versions. However, these results should be viewed with caution as only 10 data points were available for each correlation. See Table 10.

Table 10: Correlations between the original AQ and each AccAQ, for participants under 18 years old.

	AccAQ-A	AccAQ-B
Mean *(SD)*	33.36 (10.72)	30.00 (6.07)
Original AQ	$r = .92$, $p < .001$ $N = 10$	$r = .87$, $p = .001$ $N = 10$

2.3.1.7 **Exploratory Factor Analysis (EFA)**

EFA was performed on all versions of the AQ, including the original. Participants were excluded if they were under the age of 18.

Preliminary checks

Scatterplots and correlations were run to better understand the data. No inter-item correlations were greater than .9, indicating no perfect multicollinearity. As such, all items were entered into EFA.

The original AQ

To identify the optimal number of factors, PA-PCA was run which indicated that a 4-factor solution should be employed for the EFA. The EFA was run, forcing a 4-factor solution with Promax rotation to allow for correlation between the factors. The 4 factors explained 34.5% of the variance in the dataset. Both the pattern and structure matrix were viewed, these showed a similar pattern of loadings. The pattern matrix is shown in Table 11 and is colour-coded to show the fit items into their respective factors.

Table 11: Long-form original AQ factor loadings for individual items.

	Factor 1	Factor 2	Factor 3	Factor 4
Orig AQ 1	-0.46	-0.05	0.22	-0.09
Orig AQ 2	0.18	0.09	0.30	0.34
Orig AQ 3	0.01	0.01	0.58	-0.03
Orig AQ 4	0.21	0.31	-0.12	0.08
Orig AQ 5	0.15	0.45	-0.11	0.26
Orig AQ 6	-0.18	0.67	-0.03	0.32
Orig AQ 7	0.10	0.53	0.06	0.00
Orig AQ 8	-0.02	0.04	0.53	-0.16
Orig AQ 9	-0.10	0.53	0.07	0.11
Orig AQ 10	0.32	0.10	0.28	-0.14
Orig AQ 11	0.83	-0.12	0.01	-0.01
Orig AQ 12	0.07	0.42	-0.27	0.33
Orig AQ 13	0.62	-0.04	-0.12	-0.08
Orig AQ 14	0.01	-0.39	0.70	0.31
Orig AQ 15	0.55	-0.05	0.05	0.10
Orig AQ 16	0.08	0.58	-0.11	-0.06
Orig AQ 17	0.74	0.04	-0.11	-0.20
Orig AQ 18	-0.24	0.57	-0.01	-0.12
Orig AQ 19	-0.05	0.61	0.05	0.15
Orig AQ 20	0.12	0.00	-0.73	-0.08
Orig AQ 21	-0.13	-0.14	0.54	0.18
Orig AQ 22	0.75	0.02	-0.03	0.00
Orig AQ 23	0.11	0.53	-0.15	0.22
Orig AQ 24	0.25	0.09	0.15	-0.13
Orig AQ 25	0.36	0.14	0.14	0.38
Orig AQ 26	0.73	-0.07	-0.01	-0.02
Orig AQ 27	0.24	0.15	0.40	-0.10
Orig AQ 28	0.26	0.25	0.00	0.16
Orig AQ 29	-0.20	0.30	0.02	0.41
Orig AQ 30	0.10	-0.06	-0.07	0.59
Orig AQ 31	0.09	0.42	0.18	-0.17
Orig AQ 32	0.24	0.29	0.21	-0.08
Orig AQ 33	0.32	0.35	0.01	-0.15
Orig AQ 34	0.59	0.07	0.01	0.21
Orig AQ 35	-0.12	0.31	0.26	-0.06
Orig AQ 36	0.10	0.32	0.31	-0.19
Orig AQ 37	0.24	0.28	0.22	0.13
Orig AQ 38	0.87	-0.11	-0.02	-0.10
Orig AQ 39	0.04	0.44	-0.01	0.03
Orig AQ 40	0.20	0.12	0.12	-0.02
Orig AQ 41	-0.02	0.70	-0.07	0.00
Orig AQ 42	-0.11	0.35	0.41	-0.06
Orig AQ 43	0.39	-0.14	0.03	0.34
Orig AQ 44	0.76	0.05	-0.01	-0.05
Orig AQ 45	0.05	0.38	0.32	-0.04
Orig AQ 46	0.46	-0.14	0.06	0.12
Orig AQ 47	0.67	0.03	0.01	0.02
Orig AQ 48	0.25	-0.05	0.19	0.01
Orig AQ 49	-0.15	0.11	0.14	0.47
Orig AQ 50	0.13	-0.08	0.51	0.03

*Note. Darker green indicates more positive factor loadings, darker red indicates more negative factor loadings.

Component Correlation Matrix

The component correlation matrix found that some, but not all, factors were correlated, as seen in Table 12. The strongest relationship was between factors 1 and 2 while the strongest negative correlation was found between factors 3 and 4.

Table 12: Component correlation matrix for factors found in the long-form original AQ.

	Factor 2	Factor 3	Factor 4
Factor 1	.57	.36	-.12
Factor 2	-	.30	.07
Factor 3	-	-	-.21

AccAQ-A

To ascertain the number of factors that EFA should investigate, PA-PCA was run. Again, this showed that a 4-factor solution fit best. EFA was run forcing a 4-factor solution with Promax rotation to allow for correlation between the factors. The 4 factors were found to explain 41.8% of the variance. When the pattern and structure matrices were investigated to assess item loadings, they presented similar patterns of results. The pattern matrix is shown in Table 13.

Table 13: Long-form AccAQ-A factor loadings for individual items.

	Factor 1	Factor 2	Factor 3	Factor 4
AccAQ-A 1	0.84	-0.16	0.09	-0.02
AccAQ-A 2	0.03	0.19	0.07	0.51
AccAQ-A 3	-0.13	0.74	-0.10	0.08
AccAQ-A 4	0.39	0.13	-0.08	0.07
AccAQ-A 5	0.05	-0.07	0.54	-0.06
AccAQ-A 6	0.01	0.08	0.63	-0.05
AccAQ-A 7	0.25	0.08	0.01	0.24
AccAQ-A 8	-0.06	0.81	-0.09	-0.05
AccAQ-A 9	-0.09	0.13	0.59	0.03
AccAQ-A 10	0.75	-0.08	-0.03	0.08
AccAQ-A 11	0.91	-0.20	-0.17	0.10
AccAQ-A 12	0.05	-0.07	0.55	-0.01
AccAQ-A 13	0.67	-0.06	0.06	-0.14
AccAQ-A 14	-0.32	0.65	-0.09	0.28
AccAQ-A 15	0.69	-0.07	0.05	-0.11
AccAQ-A 16	0.14	0.05	0.03	0.37
AccAQ-A 17	0.67	0.04	0.15	-0.08
AccAQ-A 18	0.35	0.32	-0.05	-0.02
AccAQ-A 19	0.15	0.15	0.39	0.06
AccAQ-A 20	-0.09	0.83	0.04	-0.08
AccAQ-A 21	-0.12	0.37	0.19	-0.07
AccAQ-A 22	0.85	-0.15	-0.10	0.10
AccAQ-A 23	0.06	0.02	0.60	-0.01
AccAQ-A 24	0.38	0.20	0.12	-0.20
AccAQ-A 25	0.45	0.01	0.08	0.33
AccAQ-A 26	0.77	0.03	-0.04	0.05
AccAQ-A 27	0.26	0.61	0.08	0.02
AccAQ-A 28	0.33	0.01	0.37	0.31
AccAQ-A 29	-0.14	-0.09	0.56	0.10
AccAQ-A 30	-0.18	-0.19	0.42	0.28
AccAQ-A 31	0.42	0.42	0.02	-0.13
AccAQ-A 32	0.49	0.14	-0.04	0.13
AccAQ-A 33	0.56	0.21	-0.01	-0.05
AccAQ-A 34	0.13	-0.03	0.07	0.66
AccAQ-A 35	0.09	0.41	-0.05	0.15
AccAQ-A 36	0.36	0.46	0.17	-0.10
AccAQ-A 37	0.36	0.14	-0.07	0.26
AccAQ-A 38	0.83	-0.04	-0.04	0.00
AccAQ-A 39	0.48	0.14	-0.10	0.20
AccAQ-A 40	-0.03	0.46	0.10	0.04
AccAQ-A 41	0.21	0.01	0.43	0.07
AccAQ-A 42	0.08	0.55	-0.07	0.17
AccAQ-A 43	-0.26	0.05	0.08	0.67
AccAQ-A 44	0.78	0.01	0.04	-0.12
AccAQ-A 45	0.37	0.48	0.08	-0.15
AccAQ-A 46	0.40	-0.07	-0.20	0.38
AccAQ-A 47	0.85	-0.22	0.07	0.08
AccAQ-A 48	0.57	0.13	-0.03	0.00
AccAQ-A 49	-0.29	-0.06	0.37	0.22
AccAQ-A 50	0.06	0.46	-0.12	0.26

*Note. Darker green indicates more positive factor loadings, darker red indicates more negative factor loadings.

Component Correlation Matrix

The results of the component correlation matrix indicated that, to some extent, all factors were positively correlated, as shown in Table 14. The highest correlation was found to be between factor 1 and 2 while the weakest was between 2 and 3.

Table 14: Component correlation matrix for factors found in the long-form AccAQ-A.

	Factor 2	Factor 3	Factor 4
Factor 1	$r = .51$	$r = .19$	$r = .25$
Factor 2	-	$r = .09$	$r = .13$
Factor 3	-	-	$r = .12$

AccAQ-B

To identify the optimal number of factors, PA-PCA was run which indicated a 4-factor solution would fit best. Conducting EFA, forcing 4 factors with Promax rotation, these were shown to explain 37.5% of the variance in the data. As seen in Table 15, the pattern and structure matrices indicated a similar pattern of results.

Table 15: Long-form AccAQ-B factor loadings for individual items.

	Factor 1	Factor 2	Factor 3	Factor 4
AccAQ-B 1	0.89	-0.19	0.07	-0.12
AccAQ-B 2	-0.02	0.01	0.16	0.69
AccAQ-B 3	-0.25	0.63	-0.01	0.21
AccAQ-B 4	0.14	0.11	-0.04	0.31
AccAQ-B 5	0.07	-0.17	0.50	0.17
AccAQ-B 6	-0.19	0.11	0.57	0.01
AccAQ-B 7	0.18	0.25	0.22	0.11
AccAQ-B 8	-0.17	0.73	-0.01	-0.03
AccAQ-B 9	0.02	0.16	0.41	-0.02
AccAQ-B 10	0.67	0.09	-0.07	0.21
AccAQ-B 11	0.67	0.06	-0.14	0.31
AccAQ-B 12	0.01	-0.17	0.59	0.23
AccAQ-B 13	0.69	-0.24	-0.07	0.10
AccAQ-B 14	-0.12	0.55	-0.37	0.21
AccAQ-B 15	0.59	-0.03	0.04	-0.14
AccAQ-B 16	-0.05	0.12	0.31	0.44
AccAQ-B 17	0.87	-0.09	-0.13	-0.06
AccAQ-B 18	0.05	0.05	0.37	0.12
AccAQ-B 19	-0.05	0.14	0.64	-0.12
AccAQ-B 20	0.09	0.43	0.09	-0.07
AccAQ-B 21	-0.03	0.35	-0.05	-0.04
AccAQ-B 22	0.51	0.12	-0.05	0.34
AccAQ-B 23	-0.06	-0.08	0.67	0.06
AccAQ-B 24	0.36	0.07	0.19	-0.32
AccAQ-B 25	0.12	-0.15	0.23	0.51
AccAQ-B 26	0.81	0.04	-0.13	0.12
AccAQ-B 27	0.24	0.46	0.13	-0.05
AccAQ-B 28	-0.03	-0.12	0.62	0.20
AccAQ-B 29	-0.23	0.14	0.24	0.06
AccAQ-B 30	-0.07	-0.31	0.35	0.23
AccAQ-B 31	0.21	0.22	0.43	0.02
AccAQ-B 32	0.10	0.50	0.04	0.02
AccAQ-B 33	0.32	0.19	0.23	0.06
AccAQ-B 34	-0.10	0.26	0.20	-0.11
AccAQ-B 35	-0.17	0.50	0.02	0.21
AccAQ-B 36	0.32	0.36	0.30	-0.16
AccAQ-B 37	0.15	0.37	0.10	0.20
AccAQ-B 38	0.90	-0.01	-0.09	0.03
AccAQ-B 39	0.14	0.03	0.16	0.37
AccAQ-B 40	0.01	0.50	0.06	0.01
AccAQ-B 41	0.08	0.03	0.52	0.04
AccAQ-B 42	0.00	0.64	0.07	-0.04
AccAQ-B 43	-0.12	-0.18	0.20	0.43
AccAQ-B 44	0.87	-0.13	0.05	-0.11
AccAQ-B 45	0.25	0.30	0.19	-0.21
AccAQ-B 46	-0.03	0.30	-0.16	0.56
AccAQ-B 47	0.70	-0.02	-0.05	0.22
AccAQ-B 48	0.59	0.18	-0.03	-0.11
AccAQ-B 49	-0.07	0.03	0.21	0.05
AccAQ-B 50	0.20	0.49	-0.16	-0.07

*Note. Darker green indicates more positive factor loadings, darker red indicates more negative factor loadings.

Component Correlation Matrix

Once again, as Table 16 shows, the component correlation matrix indicated that all factors were positively correlated. The highest was found to be between factor 1 and factor 2, and the weakest between factor 3 and factor 4.

Table 16: Component correlation matrix for factors found in the AccAQ-B.

	Factor 2	Factor 3	Factor 4
Factor 1	$r = .54$	$r = .39$	$r = .32$
Factor 2	-	$r = .25$	$r = .19$
Factor 3	-	-	$r = .10$

Identifying subscales within each AQ

Using the pattern matrices for each version of the AQ, items were identified in terms of which factor they corresponded to the highest. Table 17 shows that different versions of the AQ are underpinned by different factor structures.

Table 17: Factors each item, in their respective long-form AQ version, loads to the highest according to the pattern matrices.

	Original AQ	AccAQ-A	AccAQ-B
Item 1	3	3	1
Item 2	4	4	4
Item 3	3	2	2
Item 4	2	1	4
Item 5	2	3	3
Item 6	2	3	3
Item 7	2	4	2
Item 8	3	2	2
Item 9	2	3	3
Item 10	3	1	1
Item 11	1	1	1
Item 12	2	3	3
Item 13	1	1	1
Item 14	3	2	2
Item 15	1	1	1
Item 16	2	4	4
Item 17	1	1	1
Item 18	2	2	3
Item 19	2	3	3
Item 20	1	2	2
Item 21	3	2	2
Item 22	1	1	1
Item 23	2	3	3
Item 24	1	1	1
Item 25	1	1	4
Item 26	1	1	1
Item 27	3	2	2
Item 28	1	3	3
Item 29	4	3	3
Item 30	4	3	3
Item 31	2	2	3
Item 32	2	1	2
Item 33	2	1	1
Item 34	1	4	2
Item 35	2	2	2
Item 36	3	2	2
Item 37	2	1	2
Item 38	1	1	1
Item 39	2	1	4
Item 40	1	2	2
Item 41	2	3	3
Item 42	3	2	2
Item 43	1	4	4
Item 44	1	1	1
Item 45	2	2	2
Item 46	1	4	4
Item 47	1	1	1
Item 48	1	1	1
Item 49	4	3	3
Item 50	3	2	2

Internal consistency

Cronbach's alpha was calculated for each factor of the AQ versions investigated here; results can be found in Table 18. According to Cohen (2008) scores of .7 of higher can be considered good. From the original AQ, 2 subscales fall into this range, from the AccAQ-A, there is only 1, and from the AccAQ-B, there are 3. This indicates that many of the subscales identified here are inadequate and would benefit from having items removed to improve the internal consistency. This however, will not be explored here as short-form variations for each AccAQ already exist and are being tested as part of the wider study.

Table 18: Cronbach's alpha for each factor on their respective version of the long-form AQ.

	Original Version	AccAQ-A	AccAQ-B
Factor 1	.85	.88	.82
Factor 2	.83	.66	.72
Factor 3	.65	.64	.71
Factor 4	.31	.50	.64

2.3.2 **Short-form AQ**

2.3.2.1 **Concurrent validity**

Group aged under 18 years old

For the purpose of these analyses, the under-18 and 18+ group were separated. This first set of results refers only to those under 18 years old.

Descriptive statistics, skewness, and kurtosis were calculated for the short-form original AQ, AccAQ-A, and AccAQ-B (see Table 19). Skewness and kurtosis were shown to be normal for all versions of the AQ. Scatterplots indicated a linear relationship between the original and each AccAQ, and a Pearson correlation was run between the total scores on each AQ. The results are displayed in Table 20. The correlation between the original and AccAQs were found to be significant and positive, though these are weaker than the 50-item versions.

Table 19: Descriptive statistics for the short-form AQ, AccAQ-A, and AccAQ-B, under-18 group.

	Mean *(SD)*	Min - Max	Skewness *(SE)*	Kurtosis *(SE)*
Original AQ *N* = 31	5.84 (2.12)	2 - 10	-.12 (.42)	-.691 (.82)
AccAQ-A *N* = 31	5.85 (1.97)	2 - 9	-.24 (.42)	-.69 (.82)
AccAQ-B *N* = 31	6.19 (2.02)	2 - 10	.11 (.42)	-1.29 (.82)

Table 20: Correlations between short-form original AQ and AccAQ total scores, for under-18 group.

	AccAQ-A *N* = 31	AccAQ-B *N* = 31
Original AQ	$r = .73$, $p < .001$	$r = .80$, $p < .001$
AccAQ-A	-	$r = .58$, $p < .001$

Informants

For participants under the age of 18, and where informant AQ data were available, descriptive statistics, skewness, and kurtosis were calculated; see Table 21 for details. These showed that while participants' self-rated scores were normally distributed, informant ratings were highly negatively skewed, with excess kurtosis. On the whole, this indicates informant raters scored participants toward the higher end of the scale. The relationships between the informant ratings and self-rated original, for the AccAQ-A and AccAQ-B, were visualised using scatterplots. No relationship was ascertained between the informants' AQ and the self-reported original AQ. Between the informant ratings and self-ratings on the AccAQ-A and AccAQ-B, the relationships appeared linear but weak, with the presence of an outlier in each. These checks indicated that the use of Pearson correlations would be inappropriate with this data. Instead, Spearman's Rho correlations were employed due to their ability to handle datasets with outliers present.

The results of these correlations are presented in Table 22. These indicate that the long-form informant AQ significantly correlated with the self-reported AccAQ-A and AccAQ-B, but not the original AQ. While the short-form informant AQ significantly correlated with the self-reported original and AccAQ-B, the correlation with AccAQ-A was non-significant.

Table 21: Descriptive statistics for the short-form AQ, AccAQ-A, and AccAQ-B, under-18 group, where third-party raters were available.

	Mean *(SD)*	Min - Max	Skewness *(SE)*	Kurtosis *(SE)*
Informant long (AQ-50) *N* = 27	35.20 (7.45)	9 - 48	-1.64 (.45)	5.01 (.87)
Informant short (AQ-10) *N* = 27	7.52 (1.93)	1 - 10	-1.72 (.45)	4.04 (.87)
Original AQ *N* = 26	6.15 (2.33)	2 - 10	-.21 (.46)	-.78 (.89)
AccAQ-A *N* = 27	5.76 (2.12)	2 - 9	-.22 (.45)	-1.06 (.87)
AccAQ-B *N* = 27	6.03 (2.22)	3 - 10	.06 (.45)	-1.50 (.87)

Table 22: Correlations between self-rated and third-party raters on each version of the short-form AQ, under-18 group.

	Original AQ *N* = 25	AccAQ-A *N* = 27	AccAQ-B *N* = 27
Informant long (AQ-50)	r_s = .38, p = .060	r_s = .49, p = .009	r_s = .41, p = .034
Informant short (AQ-10)	r_s = .53, p = .007	r_s = 38, p = .051	r_s = .48, p = .012

Group aged 18 years old or more

The above analyses were repeated for participants aged 18 or over. The descriptive statistics for this group were calculated alongside skewness and kurtosis. The data was shown to be not skewed with no excess kurtosis. See findings in Table 23.

Participants' original AQ total scores were plotted against their scores on the AccAQ-A and AccAQ-B; their results on these adaptions were also plotted against each other. All these relationships were found to be linear. Pearson correlations were run between these total scores, which are reported in Table 24. Correlations were all found to be significant and positive.

Table 23: Descriptive statistics for the short-form AQ, AccAQ-A, and AccAQ-B, 18+ group.

	Mean *(SD)*	Min - Max	Skewness *(SE)*	Kurtosis *(SE)*
Original AQ $N = 55$	5.91 (2.67)	1 - 10	-.19 (.32)	-1.03 (.64)
AccAQ-A $N = 53$	6.08 (2.69)	0 - 10	-.25 (.33)	-1.02 (.64)
AccAQ-B $N = 53$	6.04 (2.58)	1 - 10	.01 (.33)	-1.13 (.64)

Table 24: Correlations between short-form original AQ and AccAQ total scores, 18+ group.

	AccAQ-A	AccAQ-B
Original AQ	$r = .82$, $p < .001$, $N = 53$	$r = .89$, $p < .001$, $N = 53$
AccAQ-A	-	$r = .85$, $p < .001$, $N = 51$

Informants

Descriptive statistics for all participants aged 18 years old or more, where informants were available, were calculated and are presented in Table 25. There was evidence of a moderate skew in the long-form AQ data provided by informants, but this appeared to not be an issue when these data were plotted on their respective scatterplots. No evidence of excess kurtosis was found. Accordingly, Pearson correlations were run between informant scores and participants' self-rated scores (see Table 26). These found that the informant long-form AQ scores and participants' self-rated scores on the original AQ and AccAQ-B were significantly correlated, while the AccAQ-A was not. In respect of the informant AQ-10, the only significant correlation was with the original AQ self-rated score.

Table 25: Descriptive statistics for the short-form AQ, AccAQ-A, and AccAQ-B, 18+ group, where third-party raters were available.

	Mean *(SD)*	Min – Max	Skewness *(SE)*	Kurtosis *(SE)*
Informant long (AQ-50) *N* = 22	30.07 (9.99)	10 – 43	-.75 (.49)	-.86 (.95)
Informant short (AQ-10) *N* = 22	6.18 (2.77)	1 – 10	-.49 (.49)	-1.11 (.95)
Original AQ *N* = 22	6.36 (2.75)	1 -10	-.58 (.49)	-.51 (.95)
AccAQ-A *N* = 20	5.90 (3.09)	0 – 10	-.36 (.51)	-1.24 (.99)
AccAQ-B *N* = 21	6.40 (2.80)	1 -10	-.39 (.50)	-1.06 (.97)

Table 26: Correlations between self-rated and third-party raters on each version of the short-form AQ, 18+ group.

	Original AQ	AccAQ-A	AccAQ-B
Informant Long-form	*r* = .48, *p* = .023 *N* = 22	*r* = .43, *p* = .060 *N* = 20	*r* = .50, *p* = .020 *N* = 21
Informant AQ-10	*r* = .43, *p* = .043 *N* = 22	*r* = .39, p = .088 *N* = 20	*r* = .41, *p* = .062 *N* = 21

2.3.2.2 **Inter-item agreement**

As with the concurrent validity, participants aged 18 years or under and those aged 18 years or more were analysed separately. For the remaining analyses, these groups will be shown as distinct categories within the tables.

Kappa was calculated independently between the original and the corresponding items on AccAQ-A and AccAQ-B. Relationships were interpreted according to Landis and Koch (1977). For the under-18s, moderate agreement or better was shown by 7 items on the AccAQ-A and 6 on the AccAQ-B; for the 18+ group, moderate agreement was found in 7 items on the AccAQ-A and 8 on the AccAQ-B. The summarised results are shown in Table 27, and Kappa values for individual items are available in the ESM (Chapter 2 / Appendix Chapter 2).

Table 27: Kappa agreement between items on the short-form original AQ and each AccAQ variant.

Level of agreement (Landis & Koch, 1977)	Under-18		18+	
	AccAQ-A	AccAQ-B	AccAQ-A	AccAQ-B
No agreement (< 0 or non-significant)	3	3	2	2
Slight (0 — .20)	0	0	0	0
Fair (.21 — .40)	0	1	1	0
Moderate (.41 — .60)	5	3	2	3
Substantial (.61 — .80)	2	3	5	4
Perfect (.81–1.0)	0	0	0	1

2.3.2.3 **Internal consistency**

For each short-form AQ, Cronbach's alpha was calculated for both age groups separately to assess the internal consistency. For the under-18 group, all short-form variations of the AQ were shown to have very poor internal consistency while all versions for those aged 18 or more were found to have average internal consistency (Cohen, 2008). Results are shown in Table 28.

Table 28: Cronbach's alpha results for each short-form AQ variant.

Cronbach's alpha	Original AQ-Short	AccAQ-A-Short	AccAQ-B-Short
Under-18	.55	.51	.56
18+	.74	.75	.76

2.3.2.4 **Preference**

All ages were combined for this analysis. Participants who completed both versions of the AccAQ were asked which version they preferred; results are shown in Table 29. Removing participants who had no preference, this data was analysed using a Chi-Square test, which indicated no version of the AQ was clearly preferred over the other; $X^2(1, N = 24) = .17$, $p = .68$.

Table 29: Short-form AccAQ preferences.

	AccAQ-A	AccAQ-B	Neither
Number who prefer	11	13	1

2.3.2.5 **ROC Analysis**

ROC analysis was carried out on each short-form version of the AQ. The analysis was split between people under 18 years and people aged 18 or over, and the results are reported in separate tables. In order to predict someone's diagnostic category, those who reported being potentially autistic but who were not diagnosed were removed from the dataset to ensure a more robust analysis could take place.

Coordinates of the curve were used to identify the optimal cut-point in the data. The criteria used was a value .80 or greater on both sensitivity and specificity. If this was not possible, the point at which the sensitivity was closest to or greater than .80 was selected while keeping specificity as high as possible. The below points were chosen as these represented the best compromise. Sensitivity and specificity data for all calculated points is available in the ESM (Chapter 2 / Analysis / Output / AQ10_Part 2 & AQ10_Part 3).

For the under-18 group, the optimal cut-point was found to be 4.5 or greater. Only the sensitivity of AccAQ-B was deemed acceptable, while the sensitivity for the others, and the specificity for all, were not (Glascoe, 2005). Results are shown in Table 30.

Table 30: ROC analysis results including optimal cut-points for the under-18 group, short-form only.

	Original AQ-Short	AccAQ-A-Short	AccAQ-B-Short
Area Under Curve *(SE)*	.68 (.11), p = .161	.54 (.10), p = .769	.64, (.11), p = .240
Optimal cut-point	4.5	4.5	4.5
Sensitivity	.68	.65	.70
Specificity	.29	.25	.38

For the 18+ group, the optimal cut-point varied between AQ versions, as seen in Table 31. For the original, this was 5.5 while the AccAQ-A showed 4.5 and for the AccAQ-B, it was 4.72. It is important to note however, that when rounded to the next whole number, the cut-points for the two AccAQs would

be the same and be one point lower than the original AQ. While the sensitivity for all versions was found to be acceptable, only the specificity of the original AQ was shown to be acceptable (Glascoe, 2005).

Table 31: ROC analysis results including optimal cut-points for the 18+ group, short-form only.

	Original AQ-Short	AccAQ-A-Short	AccAQ-B-Short
Area Under Curve *(SE)*	.54 (.06), $p < .001$	.75 (.08), $p = .005$	.83 (.06), $p < .001$
Optimal cut-point	5.5	4.5	4.72
Sensitivity	.80	.88	.83
Specificity	.81	.60	.65

Short-form scores were derived from participants who completed the long-form AQ, by selecting the 10 items present on the short-form. Scores from these participants were then combined with the short-form participants to create a larger sample of short-form AQ scores. Those who had reported being potentially autistic but without a diagnosis, were omitted from this analysis.

In total, 43 participants under 18 years old were entered into the analysis. This found differing cut-points for each version, which are shown with the ROC results in Table 32. The sensitivity for all these measures was deemed acceptable, whilst the specificity results for all were unacceptable (Glascoe, 2005).

Table 32: ROC analysis results for the under-18, combined dataset.

	Original AQ-Short	AccAQ-A-Short	AccAQ-B-Short
Area Under Curve *(SE)*	.59 (0.10), $p = .348$	.52 (10), $p = .844$	.64 (.09), $p = .166$
Optimal cut-point	3.67	3.5	4.5
Sensitivity	.83	.83	.74
Specificity	.31	.13	.25

In total, 452 participants aged 18 or over were entered into the analysis, which found the same cut-points to be optimal as the original short-form only group. ROC results can be seen in Table 33. For all versions, the sensitivity and specificity were deemed acceptable (Glascoe, 2005).

Table 33: ROC analysis results for the 18+, combined dataset.

	Original AQ-Short	AccAQ-A-Short	AccAQ-B-Short
Area Under Curve *(SE)*	.91 (.02), $p < .001$	.89 (.03), $p < .001$	.91 (.02), $p < .001$
Optimal cut-point	5.5	4.5	4.72
Sensitivity	.88	.90	.86
Specificity	.81	.74	.77

2.3.2.6 **Power Analysis**

ROC Curve Analysis

Using these ROC results, power analyses were run. Each power analysis assumed a power (Beta) of .80 and a Type 1 error of .05.

Running the analysis with an area under curve (AUC) of .90, which is similar to the combined adult ROC analysis, the analysis indicates that a total of 5 control and 5 autistic participants would be needed.

Running the analysis with an AUC of .60, which is similar to that found in the under-18 studies, 97 controls and 97 autistic people would be needed (Lopez-Raton, Rodriguez-Alvarez, Suarez, & Gude Sampedro, 2014).

Cronbach's Alpha

Using the assumptions of a power (Beta).80 and Type 1 error .05, a power analysis for internal consistency was run. A 10-item questionnaire would require a total of 136 participants to detect a Cronbach's alpha of .70 (Bujang, Omar, & Baharum, 2018).

2.4 Discussion

The aims of this study were to develop a self-report version of the AQ that was more accessible to people with a mild to moderate learning disability, and which had robust psychometric properties. In order to address the former, two versions of the AQ were developed. Both of these adaptions were informed through the extensive literature and guidance available on improving the accessibility of text-based materials for people with learning disability (Buell et al., 2020; Callus & Cauchi, 2020; Fajardo et al., 2014; Hurtado et al., 2014; Poncelas & Murphy, 2007; Sutherland & Isherwood, 2016; Townsley et al., 2003; Waight & Oldreive, 2020). Despite this, some potential participants with learning disability were unable or did not want to complete the AccAQ. As highlighted in the participant recruitment section for the short-form AccAQ, seventy participants did not complete an AccAQ during this study, of whom nineteen had learning disability. This is, to some degree, expected given that past research indicates that what is considered accessible can vary from person-to-person (Townsley et al., 2003), and that systems of communication that truly enhance accessibility often need to be highly personalised (Poncelas & Murphy, 2007). The AccAQs, as adaptions of an existing, standardised questionnaire, could not be personalised to this extent. At the same time, these nineteen participants may not have completed the AccAQ due to other reasons, including simply forgetting to return their response to the centre that they were recruited from, or running out of time to complete it during the testing session. In any case, this reduction in numbers of people with learning disability prevented a separate analysis from being conducted for this group in a reliable way.

As stated, robust psychometric properties include: good reliability, internal consistency, test-retest, interrater reliability, where appropriate (Leung et al., 2012); demonstration of concurrent validity with existing measures, sensitivity of 70-80% or higher, and specificity of 80% or higher (Glascoe, 2005). By measuring the AccAQ at this standard, the results indicate that for people over the age of eighteen, either of the adapted long-form AccAQs would provide results comparable with using the full-length AQ. For both long-form AccAQs, concurrent validity was found to be good in terms of the overall measure and the subscales outlined for the original AQ (Baron-Cohen et al., 2001). Inter-item agreement between most items on the original AQ and AccAQs showed moderate agreement or better (Landis & Koch,

1977). In terms of reliability, both long-form AccAQs showed good test-retest reliability and internal consistency. Due to the small number of autistic participants completing the long-form version however, no ROC analysis was run, which does leave some question about its ability to be an effective screening tool.

Using the same criteria as outlined above (Glascoe, 2005; Leung et al., 2012), the results of the short-form are less compelling. Concurrent validity for the self-report versions was shown to be good, with both AccAQs having strong and significant correlations with the original AQ. However, the correlations between scores on the self-report and informant versions were not as strong. For most items, inter-item agreement was shown to be moderate or greater (Landis & Koch, 1977) but a higher proportion of items did not show agreement, compared to the long-form version. As for reliability, no test-retest information was collected and for the under eighteens, the short-form was found to have poor internal consistency while the eighteen or over group had only average internal consistency.

ROC analysis was carried out on the short-form AQ; first with only the participants who completed the short-form AccAQ and then including the short-form items from participants who had completed the long-form AccAQ. For the under-eighteen group, the ROC analysis found that neither sensitivity nor specificity were acceptable, yet for the eighteen or over group the sensitivity was acceptable but the specificity was not (Glascoe, 2005). When investigating the combined dataset, only the data from the eighteen or over group was shown to result in both acceptable sensitivity and specificity when cut-points of 4.5 for AccAQ-A and 4.72 for AccAQ-B were used (Glascoe, 2005).

EFA was run on the long-form original AQ and the long-form AccAQ versions. A four-factor structure was found to fit best for all, although there were some differences between the three versions regarding which items were loaded onto which factor. When the resulting factors were tested for their internal consistency, only two on the original AQ, one on the AccAQ-A, and three on the AccAQ-B, had acceptable Cronbach's alpha. Overall, the factor structure of the AQ, as used in this study, differed from the originally proposed five-subscale structure, though some of the resulting factors had questionable reliability. This result is consistent with much previous research on the factor structure of the AQ, which has found it to differ from the original five-factor structure proposed by Baron-Cohen et al. (2001). This

includes two-factor (Hoekstra et al., 2008; Murray, McKenzie, et al., 2015), three-factor (Hurst et al., 2007), four-factor (Stewart & Austin, 2009), and five-factor structures that differ from the original five subscales (Kloosterman et al., 2011; Lau et al., 2013).

In terms of the social validity of the AccAQ, the data from those who completed both versions indicated no clear preference for one version over another. As both versions of the AccAQ have similar psychometric properties when used with adults, it is recommended that potential users of the AccAQ are given a choice as to which they would prefer to complete.

The results of this study indicate that either of the two long-form AccAQs could be used in situations where a person has difficulty completing the original AQ because of poor literacy skills; using either of the adapted versions would still provide similar insight into the person's autistic traits. At the same time, due to the limited number of autistic people who completed either of the long-form AccAQ, no ROC analysis was run and consequently, its ability to discriminate between autistic and non-autistic people is unknown. Despite this, being able to complete an AccAQ may be helpful for research with people with learning disability as it provides information about a person's autistic traits, without the need for a full diagnostic assessment. As McKenzie and Murray (2015) argue, being able to conduct research without the need for a full diagnostic assessment is just one benefit of using screening tools in research. This may help address the issue of the high degree of underrepresentation of people with learning disability in autism-focussed research (Warner, Cooper, & Cusack, 2019). Though again, due to limited numbers of participants with learning disability, the true validity of the AccAQ in this group is uncertain. The evidence provided by the ROC analysis on the combined group of adults on the short-form, however, indicates that either adaption would act as a relatively effective screening tool for adults, in place of the original short-form AQ. This may be beneficial for highlighting those who require a more comprehensive diagnostic assessment (Glascoe, 2005; Public Health England, 2019).

It is worth bearing in mind some of the drawbacks of this study. While the long-form AQ component of the study had a large sample size (over 400 participants) the short-form component had a smaller sample size which impacted the power of many of the related analyses. This issue with power has likely influenced some of the results presented here, in particular the ROC analyses (Chakraborty, 2010;

Hanczar et al., 2010) and inter-item agreement (Bujang & Baharum, 2017; Tractenberg, Yumoto, Jin, & Morris, 2010), as sample size has been found to have an influence on both. The power analysis conducted for this study confirmed that the analyses on the short-form versions of the AQ were underpowered, particularly in relation to those aged under eighteen years old. Using the results of this study, a sample of at least one hundred and ninety-six participants split between autistic and non-autistic participants should be aimed for in future work to truly assess the validity of either short-form AccAQ in an under-eighteen sample.

The power analyses conducted indicated that, if the AUC and associated statistics found in the combined ROC analysis were correct, the sample was adequate. However, if the results were closer to that found in the under-eighteen studies, then there were not enough autistic participants present in the sample. It appears likely that the sample was adequate to estimate the ROC in adults, but unlikely that it is the case in the under-eighteen group. As for internal consistency, it appears that the sample is not adequate to give a reliable Cronbach's alpha statistic. In all, the results of this study should be viewed with caution on account of the relatively low number of autistic people in the sample, and the smaller number of participants who completed the short-form AccAQ.

It should also be noted that the AQ, as a self-report instrument, is designed for use with adults (Baron-Cohen et al., 2001), while both the adolescent (Baron-Cohen et al., 2006) and child versions (Auyeung et al., 2008) are completed by informants. This means that no self-report version of the AQ currently exists for those who are not adults. In the present study, the AQ versions were used as self-report tools with a sample of those aged under eighteen years old. It may be that the results reflect the fact that the tools are not suitable for use as self-report measures for this age group.

To fully validate the AccAQ as a screening instrument, two main steps must be taken in future research. A larger sample of autistic people must be included in the analyses to ensure that the cut-points found in the ROC analysis are indeed the best cut-points to detect autism. Further, the sample must include a greater number of people with learning disability, and ROC analysis should be run separately on this group to ascertain whether a different cut-point is required.

Conclusion

This chapter provides evidence that for adults, both long-form AccAQs have comparable psychometric properties to the original long-form AQ and that both of the short-form AccAQs may be effective as a screening tool for autism, though more evidence is required to confirm this. The use of the AccAQs cannot yet be recommended for anyone under sixteen years old, as the original AQ has not being validated as a self-report measure for this age group. While the AccAQ appears to be more accessible, being completed by twenty-eight people with learning disability as part of the study, it is still not accessible to many people who would benefit from being screened for autism or taking part in research, without the input of a third-party rater. Future research should explore options to make autism screening instruments suitable for this group.

3.0 Chapter 3: A systematic review of autism screening tools, used with people with learning disability.

Adapted from: Metcalfe, D., McKenzie, K., McCarty, K., & Murray, G. (2020). Screening tools for autism spectrum disorder, used with people with an intellectual disability: A systematic review. *Research in Autism Spectrum Disorders*, *74*, 101549. doi.org/10.1016/j.rasd.2020.101549

ESM: https://osf.io/e3728/ Materials found in folder titled 'Chapter 3'

3.1 Introduction

As has been discussed throughout the previous two chapters, screening tools are beneficial because they can help differentiate between people who should, and should not, be recommended for full diagnostic assessment by the relevant professionals (Allison et al., 2012; Glascoe, 2005). Screening is also useful in a research context (McKenzie & Murray, 2015). This research, relevant to autistic people with learning disability, is particularly important due to the greater challenges these people are likely to face, and the varying but generally poor outcomes that they experience (Arias et al., 2018; Ben-Itzchak & Zachor, 2007; Matson & Rivet, 2007). Despite the potential that this research holds, people with learning disability remain under-represented in published literature (Russell et al., 2019; Warner et al., 2019).

For screening tools to be effective, it is paramount that they identify those who do, and do not have the condition of interest as accurately as possible, and having good psychometric properties ensures this (Glascoe, 2005). At the same time, incorrectly classifying someone as having the condition could result in the person undergoing unnecessary further assessment, with the associated costs in time, resources, worry, and potential stigma. Incorrectly classifying someone as not having the condition may mean they miss out on assessment and support that they would otherwise have benefitted from.

It can be difficult for clinicians to identify the most appropriate screening tool to use, as a particular screening tool may not be endorsed (e.g. Centers for Disease Control and Prevention, 2020) or an endorsed tool may not be suitable for a particular purpose. In respect of the latter, as discussed in Chapter Two, the AQ (Baron-Cohen et al., 2001) while being recommended for use by NICE (2012) is not an effective tool for many people with learning disability due to it requiring a certain level of literacy. While

the adapted AccAQs are more accessible than, and in many ways equivalent to, the original AQ, for a number of people they are still inaccessible. In addition, the AccAQ is not appropriate for those aged under sixteen, given that the concurrent validity was measured against the original AQ, which itself is not designed for children. Though an AQ variant designed for younger people (Auyeung et al., 2008; Baron-Cohen et al., 2006) may be suitable for children and adolescents with learning disability, definitive evidence that these are appropriate is lacking. This means that other potential screening tools for children and adults with learning disability must be identified and assessed to see whether they may be appropriate for clinical and research use.

Recently, a systematic review was conducted to review the current evidence in terms of psychometric properties of existing questionnaires and diagnostic measures for autism in adults (see: Wigham et al., 2018). However, this only briefly covered questionnaires accessible to people with learning disability and only included articles published since 2014. As a result, there is a need for a more comprehensive review of screening measures that have been developed as, or adapted to be, screening tools suitable for people with learning disability. The present review aimed to address this need. As many researchers have used measures that were originally designed for children, with both adults and children with learning disability, the review included adult and child autism screening questionnaires. The populations investigated also comprised both children and adults with learning disability, both with and without autism. The measures are reviewed in terms of their psychometric properties, particularly the reliability, validity, and standardisation when used with people with learning disability.

The aim of the present chapter was therefore, to provide a detailed overview of the psychometric properties of autism screening tools that are available for use with adults and/or children with learning disability. This was to inform clinicians about the most appropriate screening measure available for their purpose, population, and individual being screened, as well as informing future directions for the development of autism screening tools for people with learning disability.

3.2 Search Strategy

The criteria for the literature search terms are shown in Table 34. English language papers which referred to the following in either title, abstract and/or keywords were included: autism or related term (column

1); a screening instrument (column 2); a keyword related to an instrument or scale (column 3). The terms 'learning disability,' 'intellectual disability,' and 'mental retardation' were not included as some articles may include this group but not note them in either the title, abstract, or keywords and so may be missed.

Table 34: Search strategy, one from each column must be present in the result.

	Column 1		Column 2		Column 3
OR	Autis*	*AND*	Screen*	*AND*	Instrument
OR	Asperger*	*AND*	'Red flag'	*AND*	Tool
OR	Pervasive developmental disorder	*AND*		*AND*	Detect*
OR				*AND*	Questionnaire
OR				*AND*	Quotient
OR				*AND*	Procedure
OR				*AND*	Scale
OR				*AND*	Indicato*
OR				*AND*	Identif*
OR				*AND*	Diagnos*

Literature searching was conducted in four databases: ProQuest, PsycArticles, PubMed, and Web of Science, with publication dates up to July 2019. With duplicates removed, a total of 3,068 articles were retrieved. The titles of identified articles were initially screened for relevance, then abstracts were read to determine if they were relevant to the review, paying specific attention to the participants and statistical approaches. Any article that the researcher had uncertainty about was reviewed by their supervisor, and a consensus between the two was reached about whether to retain the article. Full texts of the remaining 44 papers were read, alongside their reference sections in order to identify articles not identified in the initial search.

Articles were excluded if they were a review article, or paper which:

- Did not detail specific reliability and validity of screening tools (e.g. Reilly, 2009)
- Stated that DSM-III or earlier criteria were used in respect of diagnosis as clinicians are required to use the most up-to-date diagnostic methods (e.g. Teal & Wiebe, 1986)
- Did not compare people with both learning disability and autism, with people with only learning disability (e.g. Li et al., 2018)
- Outlined a tool which was not a screening instrument for autism per se, but instead screened for additional challenges autistic people and/or people with learning disability may face (e.g. Matson, Fodstad, & Mahan, 2009).

Other criteria also applied, for further details of inclusion criteria see Table 35.

Where it was unclear if an article satisfied these criteria, it was read in full by both the researcher and their supervisor, who then reached agreement about whether to retain it. While there was no strict cut-off point regarding date, the requirement to use at least DSM-IV diagnostic methods meant that articles published pre-1994 would not be included.

The remaining articles were read in full by the researcher and their supervisor. Each article had to evince a good quality diagnosis for both autism and learning disability. For autism, this meant diagnosis was consistent with the recommendations of NICE (2011; 2016) while for learning disability, diagnosis was in line with recommendations by the BPS (2015). Articles were also included if the participants had been recruited from a setting that was specific to people with learning disability (e.g. a learning disability hospital) or had a genetic condition that often results in learning disability, such as Down syndrome. The retained articles were scored in terms of quality of diagnosis of participants (see Table 36) and these scores were agreed upon by the researcher and their supervisor.

Ten retained articles were included in the final review. A search of reference lists identified three further papers that met the inclusion criteria, and one paper was identified from the review by Wigham et al. (2018). The final review contained 14 articles, all of which were scored for quality of diagnosis as outlined above. A full breakdown of numbers of articles identified throughout each stage of screening is

given in Figure 2. A summary of samples, including how participants were recruited and how autism and learning disability were diagnosed, can be found in Table 36.

Table 35: Inclusion and exclusion criteria for articles.

Inclusion	Exclusion
English language. Evidence of a good quality diagnosis: For autism: in line with NICE (2016, 2011) guidelines. For learning disability, either: in line with the BPS (2000) guidelines, were recruited from a specific learning disability setting, or had a genetic condition associated with learning disability.	A review paper. DSM-III or earlier diagnostic criteria. Did not compare autism with learning disability to learning disability without Autism. Outlined, investigated, or developed a tool designed to screen for additional challenges associated with autism and learning disability.

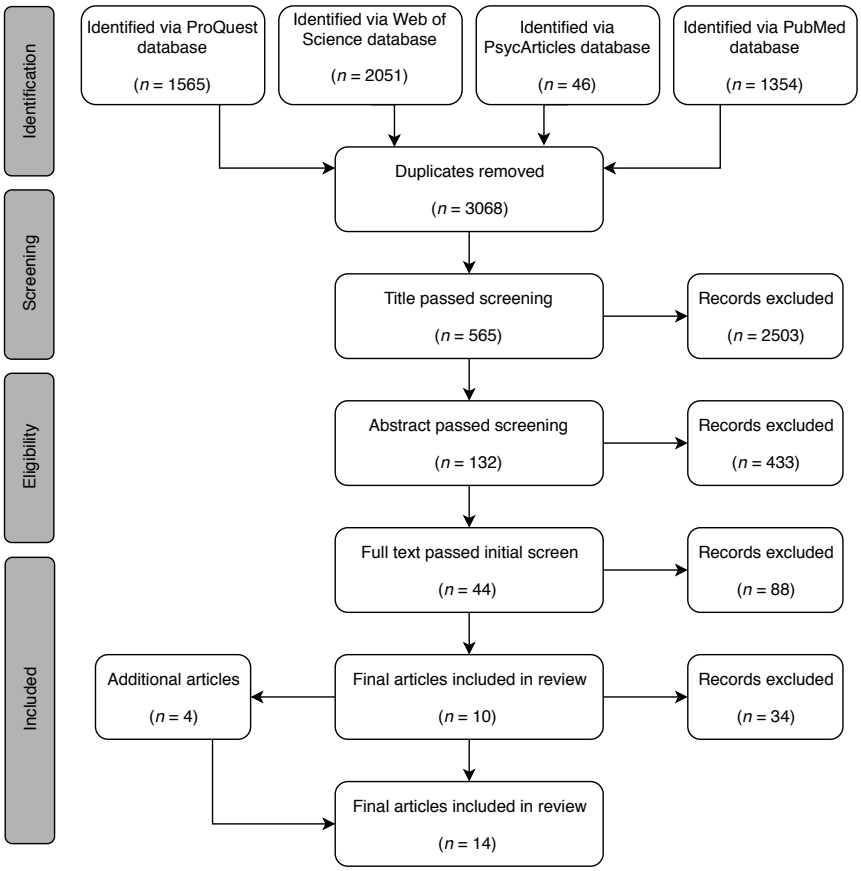


Figure 2: A PRISMA diagram showing the identification and selection procedure of included articles.

Table 36: Articles, measures used, details of sample and ratings of sample quality (organised in alphabetical order according to name of tool).

Article and measures	Participants	Autism diagnosis				Learning disability diagnosis				
		Area Under Curve (AUC), Sensitivity, and Specificity	Overall	Appropriate Clinician	ADOS-G	ADI-R	Previous diagnosis	Appropriate Clinician	IQ Test	Adaptive Functioning	Dev. History
De Bildt et al. (2003) ABC PDD-MRS	Children with learning disability Country: The Netherlands Recruited from: Facilities for children and adolescents with learning disability N = 827; N_M = 521; N_F = 306 Age < 12 N = 437 Age ≥ 12 N = 390 Profound N = 80 Severe N = 102 Moderate N = 185 Mild N = 460	ABC Groups identified according to clinical judgement. 'Autism Disorder' AUC = .76 Sensitivity = .71 Specificity = .70 'Pervasive Developmental Disorder' AUC = .75 Sensitivity = .58 Specificity = .78	5	1	1	1	1	0	1	0	0
Mutsaerts et al. (2016) ACL DiBAS-R	Adults with learning disability Country: Germany	ACL AUC = .85 Sensitivity = .58 Specificity = 78	5	1	1	1	1	0	1	0	0

	Recruited from: Department of psychiatry that specialised in learning disability N = 148 $N_{Autistic}$ = 84 M age = 38.3; SD = 12.2 Severe/profound N = 42 Moderate N = 27 Mild N = 15 $N_{LearningDis}$ = 64 M age = 34.1; SD = 11.7 Severe/profound N = 30 Moderate N = 36 Mild N = 34									
Matson, Wilkins, Boisjoli, and Smith (2008) ASD-DA	Adults and adolescents with learning disability Country: USA Recruited from: Learning disability centres N = 307; N_M = 168; N_F = 139 M age = 52; range 16 – 88 $N_{Autistic}$= 156 $N_{LearningDis}$ = 151 Profound N = 235 Severe N = 40 Moderate N = 16 Mild N = 2	Not available	4	1	0	0	1	0	1	1	0

Sappok et al. (2014) Main: DiBAS-R Extra: PDD-MRS ACL SCQ	Adults with learning disability Country: Germany Recruited from: Inpatient and outpatient services at a psychiatric clinic *N* = 219; N_M = 125; N_F = 94 *M* age = 35; *SD* = 12 $N_{Autistic}$ = 77 Severe/Profound *N* = 37 Moderate *N* = 26 Mild *N* = 14 $N_{LearningDis}$ = 142 Severe/Profound *N* = 31 Moderate *N* = 57 Mild *N* = 54	DiBAS-R AUC = .89 Sensitivity = .82 Specificity = .87	5	1	1	0	1	0	1	1	0
Heinrich, Böhm, and Sappok (2017) DiBAS-R	Adults with learning disability Country: Germany Recruited from: Inpatient and outpatient services at a psychiatric clinic *N* = 381; N_F = 161 *M* age = 40.5; *SD* = 13.4 $N_{LearningDis}$ = 289; N_F = 131	DiBAS-R AUC = .81 Sensitivity = .82 Specificity = .67 Scores given are for the whole group, scores for subgroups are available in ESM**.	5	1	1	1	1	0	0	1	0

	M age = 40.8; *SD* = 13.9 Mild/Moderate *N* = 189 Severe/Profound *N* = 100 $N_{Autistic}$ = 92; N_F = 30 *M* age = 39.6; *SD* = 12.1; Mild/Moderate *N* = 39 Severe/Profound *N* = 53									
DiGuiseppi et al. (2010) M-CHAT SCQ	Children with Down syndrome Country: USA Recruited from: A registry of birth defects *N* = 123; N_M = 80; N_F = 43 Age range 2 – 11 Children were born between 1996 and 2003, numbers per year are given within the paper $N_{Autistic}$ = 52 $N_{Non\text{-}autistic}$ = 71	M-CHAT Sensitivity = .82 Specificity = .47 SCQ Sensitivity = 1.0 Specificity =.57	5	1	1	1	1	0	0	1	0
Bergmann et al. (2015) MUSAD	Adults with learning disability Country: Germany Recruited from: Psychiatric department specialising in learning disability	Not available	6	1	1	1	1	0.5	0.5	0.5	0.5

	N = 76 *M* age = 38.3; *SD* = 11.7; Range 18 – 66 $N_{Autistic}$ = 50; N_M = 42 $N_{Non\text{-}autistic}$ = 26; N_M = 5									
Kraijer and de Bildt (2005) PDD-MRS	Large sample with learning disability Country: The Netherlands Recruited from: Previous research *N* = 1230 Age range 2 – 80 2 years *N* = 71 2 – 9 *N* = 379 10 – 19 *N* = 101 20 – 29 *N* = 168 30 – 39 *N* = 273 40 – 49 *N* = 238 50 – 80 *N* = 71 Two-year olds with all levels of learning disability *N* = 71 Persons in institutes/group homes *N* = 781 Profound learning disability *N* = 63 Persons who attend day centres *N* = 374 Persons who were attending a specialist clinic for observation and treatment *N* = 75	Compared to diagnostic status Sensitivity = .92 Specificity = .92 *Compared to ADOS-G* Sensitivity = .81 Specificity = .47	2	1	0	0	1	0	0	0	0

	Subgroups listed: Profound, severe, moderate, mild and borderline learning disability; male, female; speaking, non-speaking; age 2-9, 10-19, 20-39, 40-49, 50-80; blind/severe visual impairment, deaf/severe hearing loss; Down syndrome, and fragile X.										
Pandolfi, Magyar, and Dill (2018) PDD-MRS	Children with Down syndrome Country: USA Recruited from: A previous study of comorbidities of Down syndrome *N* = 386 Used in some analyses Sample, for tests of interest $N_{Non\text{-}autistic}$ = 38 $N_{Autistic}$ = 33 IQ Score Non-Autistic *M* = 52.38, *SD* = 14.57 Autistic *M* = 41.93, *SD* = 6.74 VABS Composite scores Non-autistic *M* = 69.65, *SD* = 9.87 Autistic *M* = 60.12, *SD* = 10.91	PDD-MRS AUC = .81 Sensitivity = .91 Specificity = .61	8	1	1	1	1	1	1	1	1

Cortes et al. (2018) PDD-MRS (Spanish version)	Adults with learning disability Country: Spain Recruited from: A wider project conducted by mental health and neurodevelopmental disorder professionals *N* = 979; *%M* = 55.7*; %F* = 44.3% *M* age = 42.4, *SD* = 13.9 Live in: Staffed residences = 52.2%; at home = 47.8% Mild = 25.5% Moderate = 28.1% Severe = 26.9% Profound = 19.5% Learning disability with genetic cause = 18.3% Down syndrome = 9%	PDD-MRS AUC = .91 Sensitivity = .70 Specificity = .91	2	1	0	0	1	0	0	0	0
Magyar, Pandolfi, and Dill (2012) SCQ	Children with Down syndrome Country: USA Recruited from: A previous autism prevalence study Group 1, exploratory factor analysis: *N* = 188; *NM* = 95; *NF* = 93 *M* age = 9.26; *SD* = 3.13	SCQ Non-verbal Cut score = 6.5 AUC = .82 Sensitivity = .82 Specificity = .68 Verbal Cut score = 10.5 AUC = .78	5	1	1	1	1	0	0.5	0.5	0

	Group 2, confirmatory factor analysis: N = 188; N_M = 101; N_F = 87 M age = 9.26; SD = 3.13 Group 3, other analyses: N = 71: $N_{Non\text{-}Autistic}$ = 38; N_M = 17; N_F = 21 M age = 7.92; SD = 3.19 M IQ = 54.38 $N_{Autistic}$ = 33; N_M = 23; N_F = 15 M age = 8.97; SD = 2.51 M IQ = 41.93; SD = 6.74	Sensitivity = .73 Specificity = .76								
Sappok, Diefenbacher, Gaul and Bölte (2015) SCQ	Adults with learning disability Country: Germany Recruited from: A university affiliated Department of Psychiatry N = 151 M age = 37.2, SD = 12.8 $N_{Autistic}$ = 83; N_M = 62 M age = 35; SD = 11 Severe/Profound N = 33 Moderate N = 39 Mild N = 11	SCQ Cut score = 15 AUC = .85 Sensitivity = .98 Specificity = .47 Scores for other cut-points are available in ESM**.	6	1	1	1	1	0	0	1	0

	$N_{LearningDis}$ = 68; N_M = 48 *M* age = 40; *SD* = 14 Severe/Profound *N* = 20 Moderate *N* = 34 Mild *N* = 14									
Derks et al. (2017) SCQ	Adults with learning disability Country: Germany, UK, USA Recruited from: Specialised learning disability and mental health services *N* = 451 (all male) Severe/profound *N* = 130 Moderate *N* = 178 Mild *N* = 143 Germany sample *N* = 261 Autistic: *M* age = 37.3; *SD* = 11.35 Non-autistic: *M* age 37.19; *SD* = 13.49 UK sample *N* = 121 Autistic: *M* age = 36.35; *SD* = 12.41 Non-autistic: *M* age = 43.64; *SD* = 12.08 USA sample *N* = 69 Autistic: *M* age = 27.62; *SD* = 5.78 Non-autistic: *M* age = 30.10; *SD* = 5.78	SCQ 'Validation Sample' Cut score = 15 AUC = .81 Sensitivity = .84 Specificity = .62 Scores for other cut-points are available in ESM**.	4	1	1	1	1	0	0	0	0

	Participants were randomly split into a training sample (N = 226) to develop a new scoring algorithm. And, a validation sample (N = 225) to validate the algorithm.										
Sappok, Brooks, Heinrich, McCarthy, and Underwood (2017)	Adults with learning disability Country: Germany, UK, USA Recruited from: Specialised learning disability and mental health services	SCQ Cut point = 13 AUC = .80 Sensitivity = .87 Specificity = .58									
SCQ	Germany sample N = 261; N_M = 181 M age = 37.3; SD = 12.3 $N_{Autistic}$ = 181 Mild N = 52 Moderate N = 118 Severe/profound N = 91 UK sample N = 121; N_M = 87 M age = 40.6; SD = 12.7 $N_{Autistic}$ = 51 Mild N = 60 Moderate N = 30 Severe/profound N = 31		5.5	1	1	1	1	0	0.5	0.5	0.5

	USA sample N = 69; N_M = 50 M age = 29.4; SD = 6.4 $N_{Autistic}$ = 21 Mild N = 31 Moderate N = 30 Severe/profound N = 8

*Note. N = Total, N_M = Male, N_F = Female, $N_{Austistic}$ = Total autistic group, $N_{LearningDis}$ = Total learning disability group, N_{Con} = Total non-autistic control group. Age in years unless specified, M = mean, SD = standard deviation. Profound, severe, moderate, mild refer to level of learning disability reported in the article. Studies published prior to the publication of DSM V (APA, 2013) are likely to have categorised severity in terms of IQ; studies after that date are likely to have categorised according to adaptive functioning. For ratings of sample quality, a score of 1 indicates that the criterion is satisfied, a score of 0 indicates that it was not satisfied or not made clear within the article, and a score of 0.5 indicates that the criterion was partially satisfied. For appropriate clinician, this means that the clinicians reported to have carried out the diagnoses of either autism or learning disability are appropriate and qualified. The ADOS-G and ADI-R columns indicate that the tools were used, as recommended by NICE (2016; 2011). For learning disability, it is possible that the sample were previously diagnosed by the institution they were recruited from or by other means. There are also three criteria which are desirable to appropriately diagnose learning disability: an IQ test; a test of adaptive functioning; a developmental history was provided, according to recommendations by the BPS 2000.

**Note. ESM found in Chapter 3 / Appendix Chapter 3.

Method of classifying results

As the review had a particular focus on the psychometric properties of the screening tools, an adapted version of the Critical Appraisal Skills Programme (CASP, 2018) checklist was used to guide the quality appraisal, along with recommendations from previous researchers about the rating of psychometric values (see Table 37). The review reports on the reliability, validity, and standardisation of a range of screening tools and Table 37 provides an overview of how reliability and validity were categorised, and the source of the classification. Results that indicate the presence of variance or invariance, or significant or non-significant differences are stated as such. Further details about psychometric properties can be found in the ESM (Chapter 3 / Appendix Chapter 3).

Table 37: Classification of psychometric properties, the rating and corresponding values.

Rating	Cronbach's alpha (α)	Pearson correlation *(r)*	Kappa	Area Under Curve (AUC)
Good	≥ .80	+/- .50-1	≥ .91	> .9
Average	.70 - .79	+/- .30-.49	.81-.90	.7 - .9
Poor	.60 - .69	+/- .10-.29	.61-80	.5 - .69
Very poor	≤ .59	+/- <.10	≤ .60	≤ .49
	(Aron, Coups, & Aron, 2013; Brace, Kemp, & Snelgar, 2016; Cohen, 2008; Kline, 1998)	(Leung et al., 2012)	(Zhu, Zeng, & Wang, 2010)*	(Leung et al., 2012)
Rating	Sensitivity	Specificity		
Good	≥ .80			
Adequate	.70 - .79	≥ .80		
Inadequate	≤ .69	< .79		
	(Glascoe, 2005)	(Glascoe, 2005)		

Note. Where a result is significant ($p < .05$) using a statistical approach which is not listed above, a rating of 'Good' will be given, otherwise a rating of 'Very poor' will be given (Leung et al., 2012). Other results may indicate the presence of variance or invariance and will be stated as such.

*Note. Originally there were five categories: 'excellent', 'good', 'worthless', 'not good' and 'very poor'. Here, 'not good' and 'worthless' have been collapsed into one category. Names of categories have been made consistent with other measures.

3.3 Results

The articles reviewed related to 8 screening tools, the majority of which were not designed specifically to screen for autism in people with learning disability but have been adapted for this purpose. The term 'learning disability' was used throughout this review to replace any terms used to indicate learning disability in the original articles, and 'autism' replaced any terms used to indicate autism in the original articles (e.g. ASD).

Background information about each screening tool was provided, followed by information about their psychometric properties *in relation to people with learning disability.* The former was sourced from general literature about the measures, while the latter was obtained from papers identified by the systematic search. While it is likely that further relevant information about some of the measures is available, only information about directly using these measures with a sample of people with learning disability has been included.

3.3.1 Screening tools evaluated only with children

Autism Behaviour Checklist (ABC: Krug, Arick, & Almond, 1980)

The ABC is an observational instrument designed to screen for autism in a large population (Bravo Oro, Navarro-Calvillo, & Esmer, 2014). The scale is a checklist of non-adaptive behaviours that reflect an individual's response to challenges in everyday life. The tool consists of 57 items (each scored 1-4) and has five subscales. The cut-point of 58 was proposed by Oswald and Volkmar (1991), meaning scores greater than 58 indicated a high chance of autism, and below less chance of autism.

Reliability and validity data

De Bildt et al. (2003) compared the ABC and the Scale of Pervasive Developmental Disorder in Mentally Retarded Persons (PDD-MRS: see below), against existing clinical classification of autism. The ABC and PDD-MRS showed agreement in 44.8% of cases, and the ABC identified 42.1% of the cases that the PDD-MRS identified, showing very poor agreement between the two. Odds ratios were significant between the ABC and both the ADI-R and clinical classification, but not when compared to

the ADOS-G. ROC analysis of the ABC, compared against clinical classification, found an average AUC.

Conclusion

The ABC shows agreement with autism classification when compared with the ADI-R and clinical classification. However, it does not show significant agreement with the ADOS-G and has low agreement when compared to the PDD-MRS. Higher scores are found in those with greater levels of intellectual impairment, which may lead to false positives in those with more severe learning disability. No reliability information was available in relation to its use with people with learning disability. Overall, caution should be exercised when using this tool.

Modified Checklist for Autism in Toddlers (M-CHAT: for development see Robins, Fein, Barton, & Green, 2001)

Designed to screen for autism in toddlers, the M-CHAT is a questionnaire completed by parents/carers, with the final version comprising 23 yes/no items (Robins, Fein, Barton, & Green, 2001). A positive screen (indicating a high likelihood of autism) is determined by either failing on any 3 items, or at least 2 of the 6 critical items.

Reliability and validity

DiGuiseppi et al. (2010) assessed the M-CHAT with children with Down syndrome. It demonstrated good sensitivity, but inadequate specificity. When combined with the Social Communication Questionnaire (SCQ), false positive results were most common in children with a hearing or a persistent visual problem.

Conclusion

While there is a great deal of research on the M-CHAT in children without learning disability, only 1 study was found relating to children with learning disability and it did not report on reliability. Odds ratios indicated that factors other than autism can affect the score of the M-CHAT. A follow-up interview is available which is designed to increase the tool's accuracy, although this was not developed with

people with learning disability in mind (Robins, Fein, & Barton, 2009). Overall, there is limited evidence that the M-CHAT is an effective screening tool for autism in people with learning disability.

3.3.2 Screening tools evaluated only with adults

Autism Checklist (ACL)

Based on International Classification of Disease – Volume 10 (ICD-10; WHO, 1990) criteria, the ACL is an observational tool which aims to identify autism in suspected cases. Each of the 3 ICD-10 domains are scored 0-4, based on the presence of each criteria. To screen positive, a person needs to have 2 points in Domain One, 1 point in Domains Two and Three, and 6 points across all 3 domains in total (Mutsaerts, Heinrich, Sterkenburg, & Sappok, 2016). No specialised training is needed to complete the checklist.

Reliability and validity data

Sappok et al. (2014) found a good correlation between the Diagnostic Behavioral Assessment for autism – Revised (DiBAS-R) total score and the ACL total score. Mutsaerts et al. (2016) later found the ACL showed significantly higher scores in autistic people compared to non-autistic people. ROC analysis found an average AUC. The ACL and DiBAS-R showed agreement in 75% of cases, with a poor Cohen's Kappa value. No information on reliability was found.

Conclusion

Information on the scale, as used with people with learning disability, is limited and further research is needed to determine if it would be a useful screening tool for autism with this group.

Diagnostic Behavioral Assessment for ASD – Revised (DiBAS-R: for development see Sappok et al., 2014)

The DiBAS-R assesses social communication and interaction in people with learning disability and can be used to detect autism. The assessment is completed by someone who knows the person well and comprises 19 items; each rated 0 to 3, which indicates how often each is true. Higher scores indicate an item is true more often. There is a maximum possible score of 57, and an overall score of 29 or more is

used as an indicator of likely autism, provided that the cut-points of subscales are met. The items are split across 2 domains in line with the DSM-V criteria of 'Social communication and interaction' and 'Stereotyped and restrictive behaviours and repetitive interests.' The maximum scores of each subscale are 36 and 21 respectively, and the cut-points are 21 and 5 respectively (Mutsaerts et al., 2016).

Reliability and validity

Sappok et al. (2014) proposes 2 factors: Social Communication and Interaction (SCI) and Stereotypy Rigidity and Sensory Abnormalities (SRS). Both factors and overall score had good internal consistency. The DiBAS-R was found to discriminate between autistic and non-autistic people, on both subscales and overall scores. A ROC analysis of the total scale showed an average (nearly good, .89) AUC. The best overall cut-point was found to be 29, requiring a score of 21 on the SCI subscale and 5 on the SRS subscale; this showed good sensitivity and adequate specificity but very poor Kappa. The DiBAS-R showed good correlations with the Social Communication Questionnaire (SCQ), PDD-MRS, and ACL. Interrater reliability was good ($r = .88$, $p < .001$, $N = 36$).

In a later study, Mutsaerts et al. (2016) found that DiBAS-R total score was significantly higher in autistic people compared to non-autistic people. A good correlation between DiBAS-R scores and ACL scores was shown, but Kappa between both measures was very poor. Those with a milder learning disability had a higher chance of a false positive result (Fishers' exact test: $\phi = .31$, $p = .017$).

Heinrich, Böhm, and Sappok (2017) found, when the DiBAS-R was assessed in the whole group, it had average AUC, adequate sensitivity but inadequate specificity. When only participants who had mild to moderate learning disability were included, AUC was again shown to be average and sensitivity to be adequate, but specificity could be considered good. When only those with severe to profound learning disability were included, AUC was shown to be poor and specificity to be very poor, yet sensitivity was considered adequate. The overall percentage accuracy of correctly identifying someone as autistic was 70.3% in the whole group; it was notably higher in those with a mild to moderate learning disability (83.3%) compared with those with a severe to profound learning disability (51.0%).

Conclusion

The DIBAS-R was designed for the detection of autism through observable social behaviour. Overall, the suitability of this measure shows mixed results, from some studies indicating good validity to others finding it to be only adequate or even inadequate on some indices. The available evidence generally suggests that it is a reliable measure, but this evidence is limited. In all, while some findings show that this tool may be an appropriate screening tool for autism in people with learning disability, more research is required before it can be recommended for wider use.

Music-based Autism Diagnostics (MUSAD: Bergmann et al., 2015)

The MUSAD was developed as a diagnostic tool built upon a music framework. It was specifically developed for adults with a lower level of functioning, including those with severe language impairments, and is completed by an observer. The test differs slightly if the person is non-verbal. The MUSAD uses music to elicit behaviours that are indicative of autism symptom severity. It encompasses 10 musical interactional situations; the final measure is a 37-item checklist scored 0 to 3, consisting of 3 factors (Social interaction; stereotypies and sensory issues; motor coordination).

Reliability and validity

The MUSAD had a good correlation with the PDD-MRS and modules 1 and 2 of the ADOS-G. It had an average correlation with the SCQ, and a poor correlation with the ABC. Interrater reliability was shown to be good between 2 raters (r = .71, 95% CI [.59, .82]) and also proved to be good between 3 raters (r = .67, 95% CI [.62, .72]). Additionally, it showed good test-retest reliability across 4 tests (r = .69) (Bergmann et al., 2015).

Conclusion

The MUSAD is the only reviewed tool that is not questionnaire-based or an informant-rated measure. It has good validity, generally showing strong relationships with other autism screening and diagnostic tools, and fair reliability. Overall, it shows potential to be an effective screening tool, but more studies are needed with people with learning disability.

3.3.3 Screening tools evaluated with both children and adults

Autism Spectrum Disorder-Diagnostic scale (ASD-DA: Matson & Minshawi, 2006)

The ASD-DA is a questionnaire completed by a third-party informant. The tool attempts to categorise people with learning disability as either autistic or not autistic, according to observable behaviour. The scale includes 31 items which are scored as either 'not different, no impairment' (0) or 'different, some impairment' (1). Level of impairment is compared with others of the same age as the target individual (Matson, Wilkins, Boisjoli, & Smith, 2008). Scores greater than or equal to 19 indicate the likely presence of autism. The scale can be split into 3 factors: Social Impairment, Communication Impairment, and Restricted Behaviour (Matson, Wilkins, & González, 2007).

Reliability and validity data

Matson et al. (2008) found the ASD-DA had a good correlation with both the DSM-IV/ICD-10 checklist and the Matson Evaluation of Social Skills for individuals with Severe Retardation (Matson, 1995), as well as an average correlation with the Socialisation domain of the VABS (Sparrow, Balla, & Cicchetti, 1984).

Conclusion

The ASD-DA was developed for use with people with learning disability and shows potential to be used as an autism screening tool for this group. The limited research indicates that it has good validity when compared with other measures, but reliability information was not available. Further research is needed.

Scale of Pervasive Developmental Disorder in Mentally Retarded Persons (PDD-MRS: for original development see Kraijer, 1990)

The PDD-MRS is a tool completed by a clinician which is specifically designed to detect Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) and 'Autism Disorder' in children with learning disability. The scale has 12 dichotomous items, indicating presence or absence of autism. Each item is weighted, and there is a maximum score of 19. In this tool, scores of 10 or more indicate a high likelihood of autism, 7 to 9 indicate a doubtful category, and 6 or less indicating a low likelihood of autism (de Bildt et al., 2003).

Reliability and validity

De Bildt et al. (2003) found that the PDD-MRS and ABC agreed for 44.8% of participants, yielding an average correlation between the two, but a very poor Kappa coefficient. Odds ratios between the PDD-MRS and the ADOS-G, ADI-R, and clinical classification were significant. AUC was average for detection of PDD and autism compared with outcomes based on the ADOS-G and clinical classification, and poor for detection of autism compared to ADI-R results. Specificity was especially high when compared to clinical classification (.92).

Kraijer and de Bildt (2005) made comparisons between people with and without PDD and a 'doubtful' group, where clinical diagnosis was not clear. Significant differences were found between people with and without PDD in all subgroups[1], apart from those with hearing deficits. When comparing those with PDD and the 'doubtful' group, significant differences were found between all subgroups, aside from those with hearing deficits and with Down syndrome. The sensitivity of the PDD-MRS was good, specificity was adequate, and the misclassification rate was low at 10.60. When compared to outcomes based on the ADOS-G, the PDD-MRS had good sensitivity but inadequate specificity.

In a later study, Sappok et al. (2014) found a good correlation between the PDD-MRS and the DiBAS-R.

Pandolfi, Magyar, and Dill (2018) investigated the performance of the PDD-MRS with a sample of children with Down syndrome and autism. Using a cut-score of 1.5 an average AUC, both good sensitivity and inadequate specificity was found.

Cortes et al. (2018) adapted the PDD-MRS for a Spanish speaking sample. The internal consistency was shown to be .71 using a Kuder-Richardson-20 test. ROC analysis was conducted with data from the whole sample and those with different levels of learning disability. Looking at the whole sample, for mild and moderate learning disability, the AUC was good, while for those with severe and profound learning disability, it was average. The sensitivity of the PDD-MRS was found to be good for the mild,

[1] Subgroups: Profound, severe, moderate, mild and borderline intellectual disability; male, female; speaking, non-speaking; age 2-9, 10-19, 20-39, 40-49, 50-80; blind/severe visual impairment, deaf/severe hearing loss; Down syndrome, and fragile X.

moderate and profound learning disability groups, while being adequate for the whole sample and severe learning disability group. Specificity was adequate for all groups, except for those with a profound learning disability where it was inadequate.

Conclusion

The PDD-MRS has undergone more research than most of the measures included in this review, most likely due to it being developed earlier than the others. It shows good validity and high sensitivity when compared with the ADOS-G, although specificity values were low. It showed good agreement with a range of other screening tools, with the exception of the ABC. Additionally, Cortes et al. (2018) showed that a translated version appears to be a useful tool in Spanish speaking samples. In all, little information regarding reliability was found, therefore more research is required. Due to the age of this tool, it should be investigated more closely to ascertain whether the items included are consistent with a more up-to-date understanding of autism.

Social Communication Questionnaire (SCQ: for original development see Berument et al., 1999)

Based upon the ADI-R, the SCQ is an informant-completed measure which can be used to screen for autism. There are various versions of the SCQ, but Sappok et al. (2015) reports that the 'Lifetime Version' is a 40-item rating scale with 2 factors (Social communication [SC]; Stereotyped behaviour and unusual interests [SBUI]). The items are scored according to whether 'abnormal' behaviour is present or not. Higher scores indicate an increased likelihood of the presence of autism, with alternative cut-off scores being used with various versions of the measure and different versions being recommended for different groups (Berument et al., 1999).

Reliability and validity

DiGuiseppi et al. (2010) found sensitivity to be good (100%) but specificity to be inadequate. Magyar, Pandolfi, and Dill (2012) assessed the performance of the SCQ with participants with Down syndrome, using a 2-factor version: SC and SBUI. Factor analysis yielded reliability coefficients of .96 for SC and .83 for SBUI. A verbal and non-verbal version of the SCQ was used in this study, depending upon ability of participants. Where possible, items common to both were analysed together. T-tests showed that SCQ

scores were significantly higher for those with autism. ROC analyses were run on both the verbal and non-verbal version of the SCQ; both had an average AUC. The non-verbal version showed good sensitivity and adequate specificity, while the verbal version showed adequate sensitivity and specificity.

Sappok et al. (2014) found an average correlation between scores on the SCQ and scores on the DiBAS-R. In a later study, Sappok, Diefenbacher, Gaul, and Bölte (2015) tested a number of cut-points of the SCQ-Current score. Using a cut-point of 15, AUC was shown to be average, with good sensitivity but below adequate specificity. Increasing cut-points to 16 and 18, classifications of sensitivity and specificity did not change, and AUC was not reported. Kappa was very poor for all 3 cut-points. The results were broadly the same when the SCQ-Lifetime score was used with cut-points of 15 and 20. Good correlations were found between SCQ Current scores and the PDD-MRS and ADOS, while an average correlation was found with the ADI-R. Using the SCQ-Lifetime score, a good correlation was found with the ADI-R, but correlations with the PDD-MRS and ADOS were not significant.

Further research by Sappok, Brooks, Heinrich, McCarthy, and Underwood (2017) looked at the SCQ across cultures. This found that scores were lower on the SCQ in females compared to males, and were also significantly affected by the country the person was recruited from. ROC analysis identified the optimum cut-point as 13, which yielded an average AUC, good sensitivity, and inadequate specificity. Three further papers examined the performance of the SCQ, however, participants were stratified by IQ or described in terms of 'delay,' rather than diagnosis of learning disability.

Derks et al. (2017) assessed the SCQ with both a training and validation sample. In most cases, results were the same for both samples. Those diagnosed with autism and learning disability scored significantly higher than those with learning disability only; a cut-point of 15 yielded an average ROC, sensitivity was good, while specificity was inadequate. When 5 SCQ items were removed (which were deemed inappropriate for participants) and a cut-point of 9 was used, an average AUC, good sensitivity, and inadequate specificity were found. Kappa showed very poor/poor agreement between final diagnostic classification and SCQ scores, in both the complete and reduced sets of items.

Conclusion

Of all the measures included in this review, the SCQ is the most widely researched. For the most part, the scale shows good concurrent validity with other autism measures (although some low Kappa values are reported) and appears to be able to identify autism well in people with learning disability. The specificity is not always adequate, but the sensitivity is consistently high. The limited information that is available about the reliability of the SCQ suggests it has good internal consistency, but no information about the test-retest or interrater reliability was identified. Overall, though this is the most researched of all measures in this review and generally shows good validity, the reliability data are still lacking.

3.4 Discussion

There are a number of instruments which have been used to aid the detection and diagnosis of autism in people with learning disability. The ABC and M-CHAT both only had psychometric information available pertaining to children. Both instruments had limited available information about validity and no reliability information, in relation to their use with people with learning disability. The ABC showed good agreement with clinician opinion but was poorer in other areas of validity, while the M-CHAT scores were influenced by factors unrelated to autism, which may lead to false positive results. The ACL, DiBAS-R, and MUSAD had limited information available, which was only in relation to use with adults with learning disability. The ASD-DA, PDD-MRS, and SCQ all had research relating to both children and adults. While the ASD-DA appeared to have good validity, the available research was extremely limited, and no information on reliability was provided. Both the PDD-MRS and SCQ were better researched. The validity of the former was generally quite good, but little information on reliability was available. The SCQ showed a mixed picture, with some studies indicating good validity, while other researchers demonstrated that specificity was found to be less than adequate. The reliability information provided was limited to internal consistency, which was good, but other types of reliability information should be investigated in future research.

Some overarching points to consider are that the majority of measures have limited research specific to people with learning disability (with the SCQ appearing to have the most research relevant to this group). Also, although many of the measures have assessed reliability with other populations (e.g. Robins, Fein,

Barton, & Green, 2001) limited research was found relating to the reliability of the measures as used with people with learning disability.

Another issue relates to the diagnosis of both autism and learning disability. A number of articles were identified as being potentially relevant for the review, but on closer examination the diagnostic processes were somewhat unclear or not sufficiently robust. For example, Matson, Wilkins, and González (2007) investigated the Autism Spectrum Disorders Diagnostic Scale for Intellectually Disabled adults, however in the article they outline that they used checklists based on DSM-IV and ICD-10 to classify individuals, rather than a full diagnostic assessment. Similarly, Arun and Chavan (2018) outline the development of the Chandigarh Autism Screening Instrument, but it is unclear whether the sample includes people who have learning disability and are autistic, or whether these are two discreet groups.

The retained articles use autism diagnostic criteria comparable to those recommended by NICE (2012), such as using the ADOS-G and ADI. That said, these recommended tools were not designed specifically for diagnosing autism in people with learning disability and can be influenced by factors relevant to this group including IQ, complex or more subtle presentation (see: Wigham et al., 2018). These authors recommend combining the ADOS-G and ADI-R for better detection of autism, and many of the reviewed articles do so (e.g. De Bildt et al., 2003). More research into the psychometric properties of autism diagnostic assessments, as used with people with learning disability, is needed to ensure they are robust, gold standard measures against which the outcomes of screening tools can be compared.

While out of the scope of the present review, a number of other conditions may affect and complicate accurate diagnosis of autism in people with learning disability (Heinrich et al., 2017; Underwood, McCarthy, Chaplin, & Bertelli, 2015; Wigham et al., 2019) and should also be considered when using screening tools. For example, autistic people with learning disability experience a higher rate of comorbid conditions, including epilepsy, schizophrenia, anxiety, and alcohol misuse, compared to those who are not, and many experience more than one (Cooper et al., 2015).

The findings of this review should be considered alongside the methodology used to identify the articles. The researcher chose not to include the term ‘learning disability’ or a synonym thereof in the search terms. This meant that the initial searches were not specific for this group. However, the strategy helped

ensure that as many potential papers as possible were included which, given the wide range of terminology that is used to refer to people with learning disability, may have been missed if specific search terms were used at the first stage of searching.

A limitation of this study relates to the variability of information that was available across the included studies, in respect of the psychometric properties being examined. No one existing system was found that summarised all of the potential statistical results and accordingly, a categorisation system had to be developed that was adapted from the Critical Appraisal Skills Programme (CASP, 2018) and through recommendations from previous researchers. While this allowed all papers to be judged by the same criteria, it is acknowledged that researchers differ in the ways they categorise results, and that classifications are, to some extent, subjective. Related to this, in some cases the statistics, which should have been classified as 'good' or 'poor' based on their values alone, would be rated the opposite way when considering the implications of the result for the performance of the measure as a screen for autism. For example, a positive correlation of .9 between a screening measure score and IQ would be rated as a 'good' correlation on the classification system alone. However, the same correlation would be rated 'poor' in terms of screening measure, as it suggests that the screening score is associated with factors other than autism i.e. intelligence. Where these conflicts have arisen, they have been noted in the text and in the ESM, but this illustrates that the statistical properties of a measure should not be considered independently of the purpose of the measure and the context within which the measurement takes place.

Finally, the review focussed on the psychometric properties of the screening measures as they related to people with learning disability. It is acknowledged that a number of the measures may, for example, show good reliability with other groups, but have no data regarding people with learning disability. This may be underestimating the reliability of measures in this respect. As Raykov (2002) notes however, good reliability underpins good validity and there is a need to explore any potential group differences when a measure is used with different populations.

In conclusion, there are several screening tools that have been used with people with learning disability. It is hoped that the information helps guide clinical decision making if professionals are considering

screening for autism in people with learning disability. The review shows that, while some tools have some evidence indicating that they may be effective at screening for autism, no one tool can be recommended based upon the evidence presented here. Therefore, any results must be treated with some caution and considered in the light of the psychometric properties of the tool, the individual with whom it is being used, and the purpose of using it. The evidence shows that there is a need for further research into all tools reviewed here, especially concerning the reliability of screening measures with people with learning disability. Additionally, in order to have confidence in the performance of any screening tool, when used with people with learning disability, researchers need to use robust methods for, and provide clear information about, the diagnosis of autism and learning disability in the participant groups.

4.0 Chapter 4: Interviews to inform the development of a new screening tool.

ESM: https://osf.io/e3728/ Materials found in folder titled 'Chapter 4'

4.1 Introduction

Being correctly identified and subsequently diagnosed as autistic can have many benefits for a person. Some of these were discussed in the introduction and include accessing interventions, such as those which build social skills (e.g. Haworth et al., 2018; Rivard et al., 2016). In addition, a diagnosis can make health professionals more attentive to conditions that autistic people commonly have, including depression and anxiety (Maloret & Sumner, 2014). It can also help those who have been diagnosed gain a greater level of self-understanding (Tan, 2018), which can help them connect with other autistic people and support organisations (Beykikhoshk, Arandjelović, Phung, Venkatesh, & Caelli, 2015; Saha & Agarwal, 2016).

Screening is one component of a comprehensive diagnostic and support system that can lead to these benefits. Understanding this wider system and the role that screening plays is important to maximise the benefits and minimise the disadvantages of said screening. The pathway that leads to a diagnosis is complex and time-consuming, historically many parents have reported struggling to get a diagnosis of autism for their children and have highlighted the need for earlier diagnosis (Midence & O'Neill, 1999). This has been echoed in more recent research. Osborne and Reed (2008) set up fifteen focus groups for parents of children with autism, and reported an overwhelming call for the processes which lead to diagnosis to be faster and easier to navigate. A survey of over one thousand families found that, on average, it took three and a half years from a concern being raised to confirmation of autism (Crane, Chester, Goddard, Henry, & Hill, 2016).

The diagnostic process can be stressful for those navigating it (Midence & O'Neill, 1999; Osborne & Reed, 2008). When interviewing parents about their experiences, Braiden, Bothwell, and Duffy (2010) found that people were generally happy with how their diagnosis was disclosed to them, that they found the information about autism clear and at an appropriate time, and were generally pleased with the communication with professionals. At the same time however, people were frustrated when they had to

fight to have their concerns listened to and stated that they did not fully understand the diagnostic process. Braiden et al. (2010) point out that stress and emotional turmoil were common amongst the families. Further research has specifically investigated parental stress in more detail, finding that certain factors can contribute to this stress, foremost of which is interaction with a large number of professionals, and professionals who appear not to be communicating or working together (Moh & Magiati, 2012). This stress will negatively impact someone's experience of the diagnostic pathway, which is important as having a positive experience has multiple benefits, including more effective coping strategies and greater acceptance of the final diagnosis (Woolley, Stein, Forrest, & Baum, 1989). This indicates that the pathway itself must be improved.

Screening for autism may improve the diagnostic experience. As discussed in the introduction chapter, screening can facilitate referral for full diagnostic assessment (Glascoe, 2005), in essence speeding up the overall diagnostic process. Further to this, McKenzie et al. (2015) found that having more information at the start of the diagnostic process reduces the time between referral and diagnosis, and one way to gain information is by using screening tools. This suggests that using screening tools to make more targeted referrals, and to provide clinicians with greater information, would make the overall diagnostic process quicker.

As discussed in Chapter Two, the AQ is the recommended screening tool for autism (NICE, 2016) but its utility is limited in people with learning disability. The previously outlined adaption of the AQ into the AccAQ shows that while they are comparable in many ways, the AccAQ is still not accessible to some, and cannot be recommended for use with anyone under the age of sixteen. In addition, Chapter Three gives an overview of other potential autism screening tools used with people with learning disability. This shows that the reliability and validity of these tools with this group is limited, indicating a clear need for more effective screening tools to be developed for use with people with learning disability.

The review of the tools also highlighted that the majority of the existing instruments rely on third-party informants to provide information about the person being screened, rather than asking those being screened to provide information. This is often because people with learning disability can have relatively poor language and communication skills (Noens & van Berckelaer-Onnes, 2004; Witecy & Penke,

2017). While informant tools can be effective for some people, the reliance on third parties means that without them, screening is often impossible. Many older people with learning disability may lack a third-party informant who knows them and their early developmental history well. While younger people with learning disability have frequent contact with family members (Forrester-Jones et al., 2006; Lippold & Burns, 2009), one in four people who are middle-aged or older do not see family members more than once in a year (Bigby, 2008). Accordingly, these people would have to rely on non-family members to provide that information. Besides family, the social circle of people with learning disability typically comprises others with learning disability or staff members (Forrester-Jones et al., 2006; Lippold & Burns, 2009). These friendships have been shown to be quite superficial and closer to acquaintances (Emerson & Hatton, 1996; Lunsky, 2006) and as such, these contacts may be unable to provide information that is needed for screening and diagnosis. This, coupled with the high staff turnover in social care settings (Butler, Simpson, Brennan, & Turner, 2010; Robertson et al., 2005) means that, for some people with learning disability, third-party based screening is either of poor quality or impossible.

The issue of person-centred care also needs to be considered. NICE (2016) recommend that people should have the opportunity to make informed decisions about their care, if they have the capacity to do so. While many people with learning disability may experience communication barriers (Noens & van Berckelaer-Onnes, 2004; Witecy & Penke, 2017), this should not preclude them from being involved in the process. Research by Wigham et al. (2008) concludes that person-centred care for people with learning disability can lead to a greater sense of empowerment and control, increased confidence, self-esteem, and overall happiness for the people being supported. By helping people to be involved in the diagnostic process, these benefits may be fostered. There has been limited research that has attempted to involve people with learning disability directly in the diagnostic process. Kenny and Stansfield (2016) attempted to use the AQ with people with learning disability, but the group was so small that no actual tests of validity or reliability could be conducted. While Chapter Two reports on adapting the AQ and making it more accessible, this accessibility is still limited, with some potential users finding themselves unable to respond to it. While some people with learning disability may be able to complete screening instruments for themselves, better quality and more accessible tools would be beneficial for this group.

In the review outlined in the previous chapter, only one tool was not an informant report tool. The MUSAD (Bergmann et al., 2015) employs an interactional musical framework in which the person is observed taking part. The investigator and the person being screened play music in time with each other and throughout a series of ten musical tasks, they are observed. Following this, the investigator describes the person's behaviour (e.g. motor coordination, imitation skills, social reciprocity). While the MUSAD appears to be quite an effective tool, its use requires a range of musical equipment and some confidence with music and singing, which many working in the field might not have. Additionally, the scoring, at least partly, appears to be based on opinion. An example of this is commenting on the person's motor skills, this indicates that the mark scheme allows for some subjectivity which, without adequate training, could be improperly applied. These factors limit the usability of the MUSAD, but it does show that observational screening is possible.

A further potentially limiting factor in practice is the accessibility and social acceptability of any screening tools. Tools that fail to meet these criteria are unlikely to be used in practice. This is borne out in past research by Richardson et al. (2017) that found in order to increase the uptake of measures in practice, they must: have high levels of usability; be practically useful, including providing information to a patient; be medically useful. Although the measures in this study were not related to the diagnosis of autism, the general findings of factors that lead to uptake and use of the instruments is pertinent to the current study. In addition to the practical aspects, a tool must have high social acceptability to have widespread use. Callahan et al. (2017) highlights that while research has been conducted on the development and review of interventions, which shows many of them to be highly effective, their uptake has not matched this. They argue that a factor which prevents widespread use is social acceptability. Citing Alberto and Troutman (2008), and Wolf's (1978) work, they define this as the consumer's satisfaction with goals, procedures, and outcomes of programs and interventions. This means that in order for the screening tool to be widely used, it also needs to satisfy and be socially acceptable to the people being screened, and those administering said tool. McNeill (2019) confirms this in empirical research on the implementation of evidence-based practice in education. McNeill found that a major influence on the use of evidence-based practice was whether it was seen to have social validity. Practices

that are deemed more socially acceptable see more frequent implementation with students and lead to pupils receiving more evidence-based support, which is in turn beneficial to them.

The aim of this research was to inform the development of a proposed observation-based screening tool. This new tool will involve participants being shown videos and their reactions to these videos being observed. The mechanisms and past research that this concept is built upon will be explored in the next chapter. This chapter aimed to inform the future development of the tool, by exploring factors that stakeholders identify as being important to the accessibility and social acceptability of screening tools.

4.2 Method

Ethical approval for the study was sought from and provided by Northumbria University's Health and Life Sciences, ethics committee. Permission to interview clinicians was provided by the NHS Trust from which they were recruited.

This research employed thematic analysis to analyse the interviews, as qualitative methods allow the exploration of beliefs, experiences and attitudes that are impossible to be gained through solely quantitative work (Pathak, Jena, & Kalra, 2013). The researcher took a semi-outsider perspective, as they are neither a clinician, teacher, nor autistic or a family member of an autistic person; they do however, have experience of the diagnostic pathway through their previous work. The approach taken was an interpretivist (subjective) approach, prioritising the experiences and perspectives of the people being interviewed, and how they view the matters being discussed. This was done with contextualisation in mind, given that participants' thoughts and feelings will all be individually influenced by the context of their lives (Flick, 2018). Emic coding was used, with no prior theories being brought to the analysis, and instead priority was given to the emerging themes which were found within the data (Peterson, 2017). A constructivist epistemological stance was taken (Braun & Clarke, 2006), given the importance of the context and the way in which experiences are socially produced during the diagnostic pathway. This method allows there to be a focus on the latent themes that are not directly discussed, but underpin the discussions taking place (Braun & Clarke, 2006).

4.2.1 **Participants**

Purposive sampling was used to recruit the following participant groups:

- Clinicians with experience in autism diagnosis and/or learning disability services
- Teachers with experience of autism and/or learning disability
- Family members of autistic people
- Autistic adults and children, with and without learning disability.

Inclusion criteria were that participants had to belong to one of the above groups and use verbal communication. Those aged 18 years or above required the capacity to consent for themselves, while those under 18 years old required parental consent. In total, 32 participants were interviewed. All participants lived in the UK and were British; 31 were white and one was mixed race. Specific ages and genders of participants are available in Table 38.

Six clinicians participated (*M* Age = 44.67, *SD* = 9.98; male = 2, female = 4). Their roles included: 3 clinical psychologists, 1 consultant psychological therapist, 1 clinical research psychological therapist and 1 psychiatrist. All had postgraduate qualifications.

Ten teachers/education professionals participated (*M* Age = 44.50, *SD* = 9.06; male = 4, female = 6). Five worked in primary schools and their roles were: class teachers/teaching assistant (*N* = 3) a deputy headteacher, and a Special Educational Needs Coordinator (SENCO). Three worked in secondary schools and their roles were a teaching assistant, a SENCO, and a learning coordinator. Two participants worked in special schools: a classroom manager and a headteacher. All had postgraduate qualifications.

Ten parents participated (*M* Age = 45.50, *SD* = 7.20; male = 3, female = 7), of whom 3 had postgraduate qualifications, 3 had attended college/university, 1 had attended an access course and 3 were school leavers. All of the children whose parents were interviewed had a diagnosis of autism. One parent was also a clinician* and 1 was autistic** and as such, these participants were reported in both groups.

In the autistic group, 4 adults took part (*M* Age = 35.25, *SD* = 8.62; male = 2, female = 2). Level of education ranged from college (*N* = 2) to PhD (*N* = 2). One adult had an additional diagnosis of learning

disability, and another had an additional diagnosis of ADHD. In addition, 5 children and adolescents with autism took part (*M* Age = 10.40, *SD* = 4.56; male = 4, female = 1).

Table 38: Demographic information of participants interviewed; asterisk denotes those found in multiple groups.

Clinicians			**Teachers & education Professionals**			**Parents**			**Autistic adults and children**		
ID	*Age*	*Gender*	*ID*	*Age*	*Gender*	*ID*	*Age*	*Gender*	*ID*	*Age*	*Gender*
7	44	F	1	50	F	6	50	F	12**	46	M
13	32	F	2	32	M	8	51	F	18	36	F
27	51	F	3	45	M	12**	46	M	20	34	F
29*	43	M	4	56	F	16	46	F	21	7	M
30	58	M	5	31	M	17	60	F	22	9	F
32	40	F	9	55	F	19	41	F	24	13	M
			10	42	M	23	44	F	26	6	M
			11	36	F	25	41 / 33	M / F	28	17	M
			14	51	F	29*	43	M	31	25	M
			15	47	F						

*Note. Same participant.

**Note. Same participant.

4.2.2 **Procedure**

Participants were invited to take part through a number of different recruitment methods. The researcher contacted organisations including schools, charities, support groups, and specific NHS departments. These organisations were provided with an information sheet detailing the purpose and aim of the study, outlining what topics would likely be discussed, how interviews would be conducted, and what the researcher planned to do with the collected information. Organisational leads were able to meet with the researcher to discuss the study in detail. Organisations who agreed provided consent and nominated an individual to act as a facilitator, to invite people to take part and direct those who may be interested in participating to the researcher.

Facilitators sent emails and letters on the researcher's behalf containing a brief outline of the study, and an internet link with both more detailed information and a form for potential participants to register their interest. There were also details of how to get in touch via telephone or email if people did not want to complete the online form. Other participants were invited to take part by adverts posted on social media and/or through word of mouth; those who were interested contacted the researcher, either by email or telephone. Interviews were conducted at the university, the participant's home, workplace, school, or over the phone. Participants were provided with information about the study through an online link, email, or in print before the interview. The information included the purpose of the study, general topics that would be discussed, how data would be collected, and what output may come from the study. This information was also provided in print at the point of the interview, where participants were also given the opportunity to ask questions before consenting to take part. Once consent was given, participants were reminded that they could stop the interview at any point, for any reason.

Before the interviews, participants provided demographic information including age, gender, job role, and details about their own or their child's diagnosis. They were then provided with information about the background and purpose of the study, including a brief overview of what autism, learning disability, screening, and screening tools are. Interviews were semi-structured using an interview schedule, which can be found in the ESM (Chapter 4 / Interview Schedule). The topics included: identifying people with autism; the screening and diagnostic pathway; implementing a screening tool. The interviewer used this schedule throughout and took notes during the discussion, in addition to recording the interviews. Interviews took between twenty and ninety minutes.

Participants were asked if they were happy to be contacted to discuss the study's findings and comment on the themes. Those who agreed provided an email address and were approached after the analysis. These participants were given a summary of the themes and the main points and were asked to comment on each.

4.2.3 **Analysis**

Interviews were audio recorded, transcribed, then analysed using the procedures recommended by Braun and Clarke (2006). All transcripts were read carefully first to obtain a good understanding of the content.

NVivo was used to highlight interesting statements and when statements on similar topics were identified, these were linked by initial codes. This was repeated for each transcript, with codes from previously analysed interviews being used where appropriate. Transcripts were analysed in the order of clinicians, education professionals, parents, then autistic people. Once all of the transcripts from one group had been analysed, the identified codes were reviewed as a whole, ensuring that all similar points were connected with a code. After, codes were linked to each other in a manner that resembled proto-themes and proto-subthemes. This process was repeated for each group. Once all of the transcripts from each group had been analysed using NVivo, the results were exported to Microsoft Word, where proto-themes were refined and developed into the themes outlined below. There were three major steps to this: the initial refinement; the collaborative refinement; the final refinement. Throughout the collaborative refinement, the researcher's supervisor supported the analysis and finalisation of the themes. All of the research team agreed upon the final themes, subthemes, and their content. These themes were shown to participants who provided feedback on each, and this feedback was incorporated into the final themes presented here.

4.3 Findings

The points made by interviewees were developed into three main themes, each containing a number of associated subthemes, as outlined below and summarised in Figure 3. Each quote that is provided to illustrate the subthemes is attributed to a participant, who is designated by their unique participant number and code for the group they belong to: C for clinician, T for teachers/education professionals, P for parents, and A for autistic people (both adults and children). Two letters indicate that the participant is a member of two groups, for example AP indicates that the participant is both an autistic person and a parent.

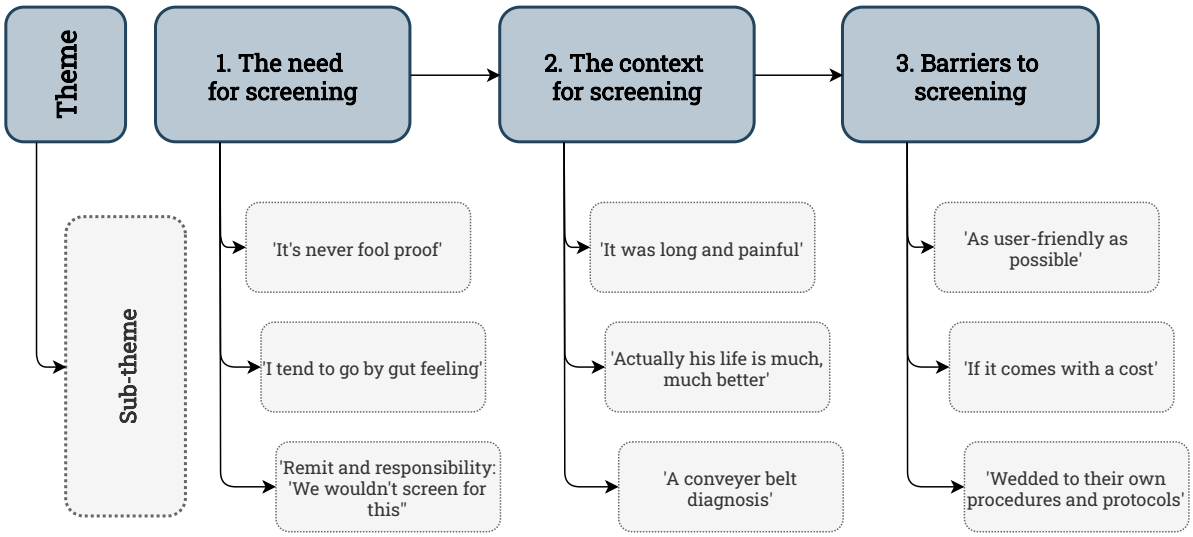


Figure 3: Themes and subthemes identified during analysis.

4.3.1 **Theme 1: The need for screening**

The first theme addressed the issue of why screening is needed and consists of three subthemes. The first, 'It's never fool proof,' identified the limited knowledge many participants had in accurately identifying traits autistic people might display. This goes onto discuss identifying people who were likely to be autistic. The second, 'I tend to go by gut feeling,' explored the limitations that the participants felt the current screening and diagnostic tools had, particularly in relation to people with learning disability. The third, 'Remit and responsibility: 'We wouldn't screen for this,'' highlighted a need for the views of concerned parties to be given weight and legitimacy, and acknowledges the role that accurate, evidence-based screening tools could have. Together, these subthemes provided a picture of the need for an accurate screening method, that could in turn, assist a range of stakeholders to clarify what autism is, identify people who are likely autistic, and provide some authority to their viewpoint when communicating with external agencies.

'It's never fool proof'

This subtheme showed the uncertainty of many participants about the traits that autistic people might display. While participants frequently described multiple traits and the clinicians often broke these down

into diagnostic generalities, the descriptions of autism were often incomplete. The descriptions acknowledge that there is a great deal of variation, but many focussed on the 'stereotypically autistic.'

> *'It's the impression that if you walked away while they were talking nothing would change, that by you being there you're being talked at and yeah you can get that with non-autistic people but there's something about the lack of basic engagement even though someone is talking at you' (C30)*
>
> *'If it's an adult, it's probably the eye contact … and so yeah certainly in the female it's eye contact. In the male, it's very much sort of, this male has very good eye contact but straight away when you talk to him you can [tell]' (P8)*
>
> *'Typical things like stimming … then sort of you can also look at sometimes repetitive patterns, repetitive behaviours that you see … repetitive behaviours that I kind of just mentioned as well, that sort of obsession and the repetitive behaviours' (P16)*
>
> *'Some of the people I see some of the behaviours are the stereotypical type behaviours so things have to be straightened' (CP29)*

To some extent, all groups felt an ability to identify autism: *'I think it's never fool proof but I think that given my experience that I would have a decent chance at spotting somebody with autism' (C7).* However, many felt uncertainty about identifying people who were likely to be autistic, especially consistently: '*Depending on the day, the time, the week, the lesson, the time in the academic year, I think possibly yes but with hesitancy' (T3)* and in different groups of people: *'When I think about kids, I think I would find it easier to spot. I think an adult, especially if they're more highly functioning, has already sort of learned skills to cope … I have two male friends who are diagnosed and one is screamingly obvious, and the other one I wouldn't have a clue' (P8).*

Many of those with more lived experience of autism, such as autistic people and their families, talked about 'getting a feeling' or what they dubbed '*Autdar.*' This is where they feel something intangible in someone else which assists them in identifying another person as likely being autistic.

> *'I wouldn't want to be like yes this person's definitely, but I think there is a thing. I think you recognise other people who are kind of like you in some way' (A18)*
>
> *'I have a friend that calls it the autdar, right, the autism radar and it's a little bit like trying to explain gaydar where you're just like, I don't know it's like, there's just there's a something. There's something that you're spotting in them that, well in my case you know what that is in me, that I'm spotting in them as well and you sort of can't pinpoint what that thing is' (A18)*

> *'There's sort of a feeling around someone with autism … It's sort of, like an array of behaviours and actions. It's sort of, it's hard to describe, it's, you meet some people and they're definitely, there's definitely something that's telling me they're on the spectrum somewhere' (P12)*

The participants also communicated a sense that it was difficult to parse out autism from other conditions, resulting in autism being used as a catch-all for many different conditions.

> *'Well I would like to think [I could spot autism] and I previously thought I would be pretty good at that but more recently I would say not necessarily … I guess I just only started to realise how different each person is' (C32)*

> *'It's now lumped onto any problem is they must be autistic, … so it's getting mixed in with other disabilities as well' (P17)*

Clinicians suggested that autism was often used as a way of opening doors to provide people with support, and that perhaps for a time there was an overenthusiasm for diagnosing autism, leading to a misunderstanding of what exactly autism is.

> *'A period when people … were over-diagnosing autism either due to misguided enthusiasm or quite honestly because it opened doors you can get more services for people' (C29)*

> *'Again this might be completely prodigious and my lack of understanding and lack of knowledge but I don't see many people who are referred into diagnostic teams for autism coming out with say a label saying you're not autistic' (C27)*

Together, this subtheme highlights a lack of clarity about what autism is, and a limited confidence in accurately identifying people who are likely autistic.

'I tend to go by gut feeling'

This subtheme addressed the past experiences of participants using more formal methods in assisting identifying people who are likely autistic.

Despite the aforementioned general lack of confidence participants felt in accurately and consistently identifying likely autistic people, the use of more formalised methods such as screening tools were rare. In particular, teachers and parents did not have experience with any autism relevant tools and lacked knowledge about them: *'We don't actually have formal screening down here other than the little sort of checklists we create ourselves*' *(T4)* and '*I think they only sort of one they used as such was the, the tools that the ed psych uses' (AP12)*.

Clinicians were more experienced with a variety of tools, but broadly lacked confidence in their usefulness.

> *'I have to say though apart from knowing what the cut off was in the AQ things like the mind in the eyes and the TOM I don't think I felt like I was fully erm. Clued up on how to interpret some of the things … so you might do something and it almost felt like I could do with somebody with a bit more knowledge just to have a look at that' (C32)*

Many also felt that diagnostic tools were too costly because of both the time required to become trained in their use, and the time needed to administer them in the clinic.

> *'[Concerning the ADOS and ADI-R] the financial costs and the time as well … it is and all the observational assessments you have go on, I've been pushing to get my ADOS training and mine's been, the message is always we're happy for you to go an observe the assessments but to actually pay for the time and your time that's where. The trust doesn't seem to be as keen' (C32)*

> *'I haven't got time to get trained in the ADOS and more to the point my trust having paid when you throw in the travel and the accommodation the best part of 800 pounds to get trained me trained in the ADI, the ADOS course is longer isn't it so you're provably doubling that, we haven't got any money [laughs]' (C29)*

> *'One of us would use a screening tool and then think about whether that means we should take it further' (C30)*

There was also a perception that the tools being used were not always appropriate: *'The whole questionnaire just kind of didn't seem to touch on my son's particular areas of difficulty' (P16)* and could fail to capture the differences amongst autistic people.

> *'I don't think screening tools even now capture the diversity of autism I suppose … I know the AQ10 is short for a reason … it did seem to just capture some people we thought were autistic but I guess sometimes we never found out for sure, but I do think it might have missed out other people who sort of erm, had differences … I just think, you know when you can't quite put your finger on it but the person is, different in the way that they communicate and the way they might approach a task or the way they interact in a group' (C32)*

This was particularly thought to be the case for people with learning disability: *'I think we don't have a, a screening measure for people with a learning disability but if it's validated, that would be helpful' (C13)*, where the interpretation of scores was considered to be an issue.

> *'I think the AQ is all very well but well, the issues with the AQ in general aside, I think for people with a learning disability, it would be a lot harder to actually understand what some of it means' (A18)*

Due to many of the existing tools being perceived as inappropriate, there was a concern that using them would lead to a failure in accurately recognising people as autistic: *'I tell you what the whole point of this and it's probably central to your question about screening tools is that people were terribly scared to get it wrong' (C30)*. As a result, many relied upon intuition rather than a properly validated tool.

> *'You need a lot of history erm and so I think for screening purposes I tend to go by gut feeling … you know if someone is wondering about autism, that wonder isn't going to go away … they reject whatever score they get on the AQ-10 because they say how sure are you and I say well not very because it isn't a very good tool' (C29)*

This subtheme indicates that little knowledge exists, outside of clinical settings, about screening tools. Within clinical settings, screening tools are used infrequently due to the lack of confidence in them.

'Remit and responsibility: 'We wouldn't screen for this''

Within the context of screening for autism, a recurring theme related to a person's remit and responsibility within their particular role. Teachers had little to no experience with autism screening and assessment tools. Moreover, many felt that their use did not fall within their remit; despite being familiar with the concept of screening and conducting screening for other conditions such as dyslexia.

> *'[Screening tools are not used] it's a medical diagnosis so we would never… we can screen, we can screen for dyslexia and that might throw something up but we wouldn't screen for this because actually its not our remit' (T3)*
>
> *'[Asked if screening tools were used] Not for autism as such but for other areas' (T11)*

While actual autism diagnosis was considered by all teachers and many clinicians to not be within their remit, they did feel part of the diagnostic pathway. In particular, as a conduit to specialist diagnostic services.

> *'You refer people who you think might have autism to the diagnostic team and they do their thing and essentially the diagnosis comes back' (C27)*
>
> *'We have a school nurse who would refer them to CYPS team and hopefully from there we'd be able to get someone from CYPS to come up and do some work with our children to help us make that diagnosis' (T14)*

While there was appreciation of the specialist skills available within the diagnostic services, there was also some consideration of whether this had undermined the confidence and skills of the participants.

'I guess this is where looking back my understanding of how the trust worked is probably not necessarily how it should have been I guess … [we] refer them to the specialist diagnostic service within the trust … instead of trying to use the skills that we had already' (C32)

'A diagnostic team that people can be referred into and over the years there's been a bit of debate as to when people would be referred to that service for diagnosis or when they would be assessed within their mental health team maybe … I think it's often dependent on whether clinicians feel confident that the team can do the assessment in the house or whether or not it might be getting scaffolding from the diagnostic team' (C13)

'I do have some concerns about the way services have gone … the way autism has become more like a service within a service' (C30)

Intertwined with this was a sense of frustration, particularly from teachers, that insights they could offer about individuals to inform the diagnostic process were not heeded by the diagnosing professionals: *'I would like to see the [teaching] professionals that work with children's viewpoint being taken a little bit more seriously … and that it's important parents have a voice' (T15)*. Having their viewpoint disregarded by professionals was also experienced by some parents.

'So it was a long, it was sort of long, sort of into the future, but it was a case of so when I started to think right there's something not right and everybody around me was like I think it's fine, I think it's new mum syndrome type thing. So once everybody, and he went off to a health visitor and kind of said he's not pointing and gesturing and things like that and she was like I know what you're trying to say to me but I don't, she said, I don't think that' (P16)

Conversely, clinicians often doubted the capabilities of education professionals.

'What you need is screening plus a kind of informed clinical judgement or an informed educational judgement, but people struggle … if you look, people, non-clinicians struggle making, they'll say oh, so the SENCO will do the screening and oh no we won't chase that up, they didn't score on the screening. They sometimes struggle to know when to apply the rules and when not to' (C7)

In this context, using an evidence-based screening tool was anticipated as one method of giving weight to the views of non-experts.

'I like the idea that it's something tangible and you know that's something that we don't have at the moment, we don't, we can put wording into a report that, you know, talks about some difficulties with communication but if it also came with a score or whatever it was then maybe it carries some weight' (T5)

'I say any any any evidence towards which comes with the you know gold embellished stamp to say its legitimate, yeah great' (T2)

Indeed, a wide range of people from different service settings were identified as being appropriate users of the screening tool. This included those working in community services: *'It'd be useful to have people like community nurses or speech and language therapy to be trained so that they could do that screening' (C13)*; in both adult mental health services: *'I would really like people to screen routinely in at least secondary mental health services because it's, it's a really obvious thing' (A20);* and child and adolescent mental health services (CAMHS).

> *'I'm not a CAMHS practitioner you see I don't usually but I have occasionally changed a diagnosis that the CAMHS team have said no and I've changed it to a yes, but usually that's because some new evidence has come to light as the person's gone on living' (C29)*

In particular, participants suggested that new screening tools would be useful for education professionals.

> *'I think it's the sort of thing that might be particularly useful for school aged kids … if you're in school and teachers are starting to wonder is he, what's going on here so that might be a very easy way of just doing a first dab' (C30)*

> *'The more complicated it gets, probably you're leaning towards a teacher or a SENCO or whatever [if it is simple] we would probably have a TA or a HLTA who would be the designated expert if you like' (T5)*

Overall, people sought screening tools wherever they are required, which may include being used by concerned parents.

> *'Because to me that as a first stage screening tool, the more accessible it is, the better. So we could be talking about a GP you know in the doctors surgery. We could be talking about you know I don't think that should be down the CAMHS route personally because why wait to get it there. That could be in a school, do you know what I mean, that could be made available to parents even … to be honest, if you're a parent who's worried about their child or a teacher who is worried about a pupil, you want something to grab onto immediately, you want something accessible immediately, and if that were to be a first stage screening tool to be accessible as possible, I think would be good' (P16)*

> *'I think it needs to go at different points doesn't it, … it would be good if it was available as required' (C27)*

> *'I do, I do like the, the accessibility of things like the AQ. I think it would be cos, I think parents, you know, there's a lot of parents who are kind of in denial but there's also lots of parents who do pick up a lot sooner than anyone else that there's something going on … Because their kids trying so hard to fit in at school that the teachers would never have a clue. So I think it would be useful for, I think the more screening tools are accessible to the public in general is, is the better really' (A20)*

In summary, the first theme explored showed that on the whole, people have an understanding and awareness of what autism is, but for many this was incomplete and based upon stereotypical behaviours. There was often an uncertainty around whether they would be able to identify autistic people, with many who have a lived experience of autism relying on a gut feeling to spot it. With this level of uncertainty in mind, many people had little to no experience of using evidence-based screening tools, and many who did, did not trust the results and instead, opted to use their own intuition to identify autism. Although there was disagreement about what is within each person's remit, many saw themselves as being a part of the diagnostic process for autism. There was a frequent expression of frustration between the clinicians and education professionals, with education professionals feeling ignored, and clinicians expressing a distrust of education professionals' opinions.

4.3.2 **Theme 2: The context for screening**

The second theme explored the diagnostic pathway for autism and how screening fits within these existing systems. It had three subthemes: The first, 'It was long and it was painful,' highlighted the importance of early identification of autism, and the role that screening can play in this. The second, 'Actually his life is much, much better,' drew attention to the fact that screening, and subsequent diagnosis, need to be of benefit to the person to be worthwhile undertaking. The third, 'A conveyer belt of diagnoses,' explored participants' views about post-diagnostic support. Together, these outlined the context that a new screening tool would fit into.

'It was long and it was painful'

The first subtheme emphasised the benefits of early autism identification, the role that screening could play within the facilitation of this, and the frustration felt by many people at a delayed diagnosis. Early autism diagnosis was seen as important because it allows support to be put in place for the person and their families as soon as possible, with the aim of minimising distress and maximising the person's life chances.

> *'I think the earlier people are diagnosed, the better … because I, I really believe that if you know what is going on and driving the behaviours and the way that the child or young person's thinking, it really helps people at home and at school and wherever else' (P19)*

'If it was diagnosed early, I don't think people would suffer as much as they do … I think people have suffered through non-diagnosis and I don't just mean people with special needs, I mean people with autism who have gone to school and wondered why they are not as clever or can't write like the person beside them and I think that's very, very sad' (P17)

Often, parents reported being the first person to identify difficulties that their children were experiencing at an early age.

'That was me [who first raised questions] … he was about. He wasn't more than 18 months … he used to go to a child minder and then she got to the point she couldn't cope with him, so he ended up going to nursery … he was displaying the usual obsessive behaviours and things' (P12)

In a similar vein, family members raising concerns about a child's difficulties and highlighting that it is possible they are autistic, was particularly helpful when the presentation was complex, unusual, and subtle.

'My dad did [first spotted my autism] because he has autism because it's easier for him to spot it cos he's smart and I'm smart so it's easier to see as if you weren't like I am it's harder to spot like if I have autism or not cos they would just say well you're smart so' (A22)

'He was given his ADHD diagnosis after a lot of to-ing and fro-ng we decided yes we would go for medication just to see if it would help him … [the doctor]'s going through the various health questions but she's also talking with him and things like that and [at the end] she said I'm not going to give you the medication … because I'd like to do some more tests because I strongly believe [son] is autistic' (AP12)

Despite the importance of early identification, many participants experienced challenges with even beginning the process because of long waiting times and lists: *'You just kind of acknowledge it and then you sit in a waiting list for a year.' (A18).* This was also acknowledged by clinicians: *'The adult autism diagnostic service…have horrendously long waiting list.' (C7)*. Even after people have entered the system, a great deal of time could be spent being referred between services…

'It can take 2 years to get a diagnosis so if they're referred to a paediatrician you have to wait for an appointment then they might refer them to CAMHS or CYPS … or they might say we're going to observe them more, 6 more months and come back' (T4)

…or waiting for existing information to be integrated into the assessment process.

'I'd like to see CAMHS pick up [forms] quicker and I'd like to see that pathway shortened in terms of the length of time it's going to take to reach a diagnostic formulation' (T10)

In all, many reported the diagnostic process as being long and difficult.

> *'It was, it was long and it was painful I would say and we'd been through a lot of grief before we even got to the point … of being at specialist services, you know, cos you get, you get a lot of negative things said to you and when you don't really know why those things are happening, it's quite hard' (P19)*

A number of reasons were identified for the length of time the diagnostic process could take. This included the complexity of the person's presentation: '*I think there's a group of people who you can diagnose and assess in about 20 seconds … and then there's a group who it doesn't matter how much assessment you've done [laughs] it's still unclear' (C30)*. This is particularly important when considering people with limited verbal ability and who require someone to support them through the assessment process.

> *'I think it would be really challenging if, cos I didn't, I didn't have my parents involved in my diagnosis so it was, you know, a clinical interview so like if you, for someone who didn't have much in the way of verbal ability, that would be really tricky' (A20)*
>
> *'If somebody was of a higher functioning individual who could just be interviewed themselves, they would just see them. When somebody was less able, I suppose they would work with the family' (C13)*
>
> *'I wouldn't automatically go to further assessment unless it was through discussion with the person if they could engage with the discussion or would their family or the carers or the care team around them' (C27)*

This engagement with additional persons, particularly parents, throughout the diagnostic process was seen as vital and beneficial: *'Obviously talk to parents, parents need to be involved with every step' (T1)* and *'I never ever do anything without having a conversation with the parents because it's so important that they're part of the process' (T15)*. This engagement could however, also lengthen the time that assessment took.

> *'Keeping people in the loop takes time and it's time you're not spent doing something else … if I'm keeping someone in the loop I could be spending that time diagnosing that person over there quicker so that's a bit tense really because I think there is a lot of keeping people in the loop because I think if families are stressed an unhappy I think their ability to look after someone is diminished … but with some families that can become endless' (C29)*

This was particularly the case if the parents were reluctant to acknowledge the challenges which their children were faced with.

> *'Bit of a softly softly approach with parents. Some parents can be very accepting but other parents can be very you know, its developmental, he's an august birthday, it could be this, it could be that' (T5)*

'There was [name of child] last year whose parents just did not want to entertain it, and [name of child] because his mam just didn't want to, I don't know, be confronted if you like by her child [being] different' (T9)

'You can tell that [some children] have, that they're on the spectrum like I had a little girl last year but mam and dad just refused to discuss it and did say, you know, she's quirky, what difference will a diagnosis make?' (T9)

Furthermore, the assessment process for autism itself is time-consuming, and clinicians must balance the needs and priorities of different groups.

'To do an ADI it takes me … the best part of two hours and if you compare that to my outpatient work sometimes people with severe mental health problems with lots of risk um lots of morbidity illnesses affecting their lives suicide risk homicide risk I can manage them in an outpatient appointment in 20 minutes … so I can see six people like that in the time it takes me to do an ADI for one person who is not risky … who is not suffering severe morbidity but the professional part of me is conflicted, yes it needs doing but is it a priority but realistically how many can I actually do so if I do them at the rate that would keep the families as content as I would like them to be, I would neglect a different group of patients' (C29)

This subtheme gave an insight into the length and difficulty of obtaining an autism diagnosis for everyone involved in the process.

'Actually his life is much, much better'

Here, the views of participants in relation to the relative costs and benefits of screening and subsequent diagnosis were explored. There was a view, shared by many participants, that what comes after the diagnosis in terms of support, and how this impacts the person who may be diagnosed, was important. It was stressed that this should be considered before embarking down the diagnostic pathway, by both parents: *'Do they want to give their child a label? Is it going to benefit them?' (P6)* and diagnosing clinicians: *'I ramble on about my hobby horse you know just the general caveat of when we do these things are they helpful to people' (C27)*. Participants highlighted that beginning the diagnostic process for autism, was rarely undertaken lightly, and was often precipitated by difficulties that the person was experiencing.

'Usually when people are coming into services for a diagnosis it's triggered by some sort of difficulty there's not that many people who are just curious and just want to see whether I've got this or not and then just crack on' (C27)

There was also a recognition that a diagnosis alone would not solve whatever difficulty the person was facing at that time.

> *'[A diagnosis is] not going to necessarily going to solve the marital problems or the depression that the person has been living with for a while so it's making sure that after diagnosis there's those links back into the services that might be needed for the person' (C27)*

Many felt however, that the diagnostic process had the potential to lead to numerous benefits, such as the newly diagnosed autistic person having a greater understanding about themselves…

> *'Sometimes someone with autism doesn't necessarily need sort of a bank of experts behind them to sort of help them on, sometimes they just need the understanding of why they do what they do … so they can learn to moderate it better you know' (AP12)*

…and others having a more complete understanding of them.

> *'Generally when you have a diagnosis of something it'll often come with ways that you can help or support the student. So it'll be specific to them and what their difficulties … so staff can put things into practice in the classroom to try and overcome the things that they face' (T11)*

This was seen as particularly important in the case of people with learning disability, where the person may not always be able to clearly articulate their needs and difficulties and thus, they may rely more heavily on the support and understanding of others.

> *'I come across situations where people haven't recognised that someone has autism in addition to a learning disability and I think that recognition and understanding … can be very, very valuable at times' (C7)*

For parents of children with autism, an additional benefit was that they could offer an explanation to others about the difficulties their child was experiencing.

> *'You're struggling to explain what difficulties your child has, well he has this he has this corpus collosum [it] means nothing to nobody, and that syndrome would have meant nothing to anybody and to just say autism would have been so much easier … to explain what his condition is with school, with everyone … and to be to, I guess manage him better, now knowing that's what he had' (P6)*

While recognising that the diagnostic process can be difficult and time-consuming, for many these drawbacks were outweighed by the eventual benefits and payoffs.

> *'We struggled for that period of time and had to do lots of paperwork and all that malarkey but the result is that actually his life is much, much better. His level of education has improved because he's able to access it. His future is a lot brighter and that's, it's worth it' (AP12)*

Screening was viewed as playing a part in this process, with increased autism identification being seen as a potential means of reducing stigma.

> *'I think the more people are screened and the more people are diagnosed, and the more people are like actually this is everywhere, the more there's going to be understanding around it and I think the more that there's understanding, hopefully the less that there's stigma' (A18)*
>
> *'Could like diagnose more people since it's much more widespread than we know' (A28)*

'A conveyer belt of diagnosis'

This subtheme explored the views of participants about the support that is provided after the screening and diagnostic processes are completed. A common concern, particularly for adults, was that the support offered post-diagnosis was limited, thereby restricting the potential benefits of screening and diagnosis. Numerous participants felt abandoned after receiving their own or their child's diagnosis: '*... and there you go, and literally there was a follow-up appointment and that was it. You're sort of cast out into the world by yourself' (P12)*. This was at a time when they were perhaps most in need of guidance and support.

> *'Obviously I'd like there to be a post-diagnostic service ... there's not much in the way of post-diagnostic support. So I did have one post-diagnostic meeting like 6 weeks after but it still hasn't really sunk in at that point ... it's like here is a lifelong neurological condition, bye' (A20)*

Indeed, some participants felt that individuals went through the long diagnostic process, and then at the end of it, did not experience any real change to the support that they received.

> *'If they've gone through a conveyer belt of diagnoses and then come out at the other end of it and because you've got that diagnosis that's kind of the generic advice that we give all schools ... much of which we would do already' (T10)*

This theme highlighted the main features around the screening and assessment process. Firstly, that difficulties may be experienced with the screening and assessment process, which were indicated through the first subtheme, 'It was long and it was painful.' Secondly, although a timely diagnosis was seen as beneficial, people encountered long waits before they were even able to access the diagnostic systems, and once accessed, further delays and frustrations made it difficult for many involved. Thirdly, the subtheme, 'Actually his life is much, much better,' highlighted issues that concerned parties, such as professionals, parents, and autistic people, should take into account when considering undergoing

screening, and potential subsequent diagnosis. Diagnosis alone was recognised as not being a panacea, instead it had the potential to lead to an improved understanding of autistic people and their needs. Screening was viewed as one aspect of this process which could lead to better recognition of autism, and in turn potentially reduce stigma. However, the identification of autism in the absence of any benefit to the person was not considered to be valuable. This issue of post-diagnostic support was discussed in the final subtheme, 'A conveyer belt of diagnosis,' where some participants described feeling unsupported and abandoned at the end of the diagnostic pathway. Others felt the diagnosis changed very little for them in terms of support, leading to a feeling that the formal diagnosis was not worthwhile. In all, this theme describes the various issues relating to the screening and diagnostic pathway as it currently stands and that while a diagnosis can help many, not everyone benefits from it.

4.3.3 Theme 3: Barriers to screening

The final theme explored the potential barriers that participants identified when introducing a new tool to screen for autism. It contained three subthemes: the first, 'As user-friendly as possible,' highlighted the need for any new tool to have good psychometric properties and accessibility. The second, 'If it comes with a cost,' explored practical barriers, such as budgets and staff training. The third, 'Wedded to their own procedures and protocols,' identified psychological barriers, including resistance to change.

'As user-friendly as possible'

This subtheme reflected the need for any new screening tool to be fit for purpose, effective, have good psychometric properties, and be acceptable to users.

In terms of psychometric properties, while clinicians used specific psychometric language such as false positives and validity: *'What's your cut-off point? What's the risk of false positives? … who's going to ensure validity?' (C7)*, the same idea was conveyed by non-clinicians: *'It would need some extensive testing and…like reliability taken into fact does it actively reflect the autism' (A28)*. Participants also suggested the benefits of triangulating the results from the screening test with other sources of information: *'I guess I would be slightly reluctant for it to be based purely on observation because I don't know how much variation there is in how people react to things anyway' (A18)*. On top of this,

there was an awareness of the need to account for factors other than autism that may influence the results. This included age, developmental level: *'Make it like age appropriate and obviously disability appropriate' (P25)*, learning disability…

> *'Often people with a learning disability have got poor theory of mind and poor empathy as a result of the learning disability but that might not be around sort of having autism. So my first thought would be would it be so sensitive to teasing apart people typically, typically intellectually impaired people versus autistic people' (C13),*

…and individual and environmental factors.

> *'I don't know if it comes down to… obviously a lot of things are going to affect your reaction. So should it be at a certain time of day, should it be morning, afternoon. Should they have just eaten, should they have not' (T5)*

Many participants discussed the importance of the tool being accessible, irrespective of the level of ability of the person being screened: *'I don't know what kind of formats people prefer to respond in if they, in terms of intellectual disability 'cos most people you just ask questions verbally so you'd have to… make it as user-friendly as possible' (A18)*. This was seen as particularly important for those who do not have someone who can advocate on their behalf: *'I think it's probably a bit of an assumption on everybody's part that everybody with an intellectual disability has someone who is willing and able to actually do that and advocate for them' (A18*). Parallel to this, the need for clear, straightforward instructions was stressed: *'So clear instructions yes because some of you know, you're going to get a whole range of difficulties, aren't you?' (P6)*, with complexity being identified as a barrier to any tool being used: *'It needs to be easy to use because anything at all really complicated or requires a manual doesn't get used' (T10)*.

Interviewees crucially recognised and drew attention to the important distinction between an accessible tool and one that might appear patronising to users.

> *'Yeah, I think again it goes back to the usual things a lot of people with learning disabilities get annoyed if you use picture that are, somebody can think you're treating them like a child' (C32)*
>
> *'I think for older children say teenagers I think you're gonna get those who think it's patronising.' (P25)*

In terms of general accessibility, the use of videos displayed via computer screens was viewed positively.

'If there's any reading or anything like that involved, that can obviously be a problem because they're not always able to do that. So I think you know like you say little video clips or whatever is ideal because that's accessible right across the board. So there's no as you say barrier for them.' (T9)

'It seems quite a, a simple way of measuring somebody's response. It doesn't necessarily rely on their language ability or their cultural ability or their families sort of learning … this is a human sort of response so I think that's quite a levelling thing probably' (P19)

As well as emphasising the clear instructions and accessibility required for any tool, there was an insistence on ensuring it was as engaging as realistically possible. A potential barrier to this engagement was that the tool might take too long to complete, although there was a lack of consensus about how long was too long.

'It's got to be something that's good, quick and efficient … yeah you've got to do it in a time of, let's say a lesson, so therefore it's got to be done within an hour.' (T3)

'You know, an hour would probably not work' (T15)

Many participants considered that people might engage well with videos delivered via a screen.

'I think he would have engaged with something like that better than like this, a face to face interview with another person cos he's comfortable with a screen' (P16)

'I think they'll [be] fine, if you make it fun which I'm assuming you're going to … I think they'll enjoy it … It's on a computer, they all love to do that. I'd just say make it fun and I think they'll enjoy it' (P6)

'I love watching videos … I've never thought they would help spot autism like I've never thought that' (A22)

'I remember in year three or something and someone asked a few questions [about videos] like do you think they are like how do you think they are acting, one of them picked up a giant huge ball and threw it down and went, stomped off, I was like well he's not angry I was rubbish at it … I loved it' (A22)

Nevertheless, it was recognised that such an approach would not be engaging for everyone: *'Some of our clients wouldn't watch a video … some of our adult clients they're just like what's this, no and that will be the end of it' (C7)*. While others considered that some people may find being observed somewhat anxiety-provoking.

'Would you mind other people watching you while you watch the videos] yes … I wouldn't like it … because there's someone watching you … [what about being watched through a video camera] then I would like that [is that because you feel like they're not looking] yes' (A24)

'I wouldn't like any other kids watching from afar or anything, like if they were coming in from break, like if they were looking in from afar and decided to take advantage of it' (A22)

This subtheme drew attention to the many issues that may present themselves during the development and subsequent use of a new screening tool. Primarily, this concerned the accessibility, as participants felt it was vital to make any proposed tool accessible to both the people completing it and those using it, in order for it to be used.

'If it comes with a cost'

In this subtheme, participants identified a number of potential practical barriers that would need to be overcome in order to successfully introduce a new screening tool. An important consideration was the financial cost, and the need to ensure that the benefit of the tool outweighed the cost of it, due to the constrained budgets of many of the services in the area.

> *'If it comes with a cost attached we won't buy it, there's no money or you'd have to demonstrate efficacy, really high levels to make it worthwhile for any diagnostic service to use it.' (C7)*

> *'I don't really hear about people using screening tools which aren't just bits of paper because like, you know, budgets … they're always looking for cost-effective things so I guess if you can argue that it's very cost-effective then they might' (A20)*

When considering the development of a new tool, participants stressed that it is important to look at the barriers to its use that might arise and then do what is possible to mitigate against them. A related resource consideration was around training. In addition to keeping costs down, participants felt that the training must ensure those using the screening tool had a clear understanding of its purpose, how to use it, and how to interpret its results.

> *'I think it'll be making sure people are properly trained and understanding' (C27)*

> *'I think it's making sure that there's proper training in place for teachers if they are going to implement this, because I think if I was to just pick it up and read a book I think I'd be like oh have I read it right kind of thing … and sometimes it's nice if you did training if you did staff meetings or inset days' (T14)*

Once trained, the next concern was about people having the time to use the tool in services that are already over-stretched.

> *'As a screening tool just be careful that it's not something that's completely out of the blue and additional burdensome there are health screening tools so think about ways in which something like this can be brought into already existing screening' (C30)*

'I suppose schools themselves might be reluctant if it's one more thing that they have to do or something because obviously there's so much stuff that year on' (T11)

While relatively short, this subtheme exemplified barriers that would likely hamper a new screening tool in use; many of which were practical and could be addressed during the development of the tool.

'Wedded to their own procedures and protocols'

This subtheme explored psychological barriers to the introduction of a new screening tool, such as resistance to change: *'The services in general are a little bit, a little bit tricky at, at… it's quite hard to get people to do things differently' (A20).* This was seen as being particularly the case if people already used established tools: *'If people are already using certain measures then sometimes it's hard for them to, to change' (C13)*. The NHS was identified by both clinicians and autistic adults as having set structures and patterns which were also likely to inhibit change.

'Other than the NHS not liking to do things differently like, yeah. Them with their, you know, rigid inherent routines and stuff. I don't know really … trying to get people to do new things is hard' (A20)

'[Where a tool would fit] is an issue because if you think about … the health pathway to get into our pathway you've got to have big problems … they're funny buggers I think people get very wedded to their own procedures and protocols so if you're not coming up with something that is solving their problem so if you're just coming up with a new idea, then you find that it may appear more effortful than useful does that make sense' (C30)

These psychological barriers were also thought to exist in both parents and teachers.

'I think your barriers will be … do they have a SENCO and how open are, are the schools to saying yeah look we've got a problem and you know … [the school's] still got some teachers who are going down, going down an old-fashioned path of not being open to it' (P8)

'I think some people are naturally not very open to, to trying sort of new things are they so I think that's just a case of making sure people understand what it is … parents more so I think' (P23)

A key challenge was seen to be demonstrating that a new screening tool would have sufficient benefits to overcome any resistance to change.

'I suppose trying to introduce anything new you have to convince people of the merits of it so that's a barrier, you know you need to be showing them that it's time efficient and reliable and gonna give you better results than something else or at least as reliable' (C27)

A highly desired feature of any new screening tool was that those being screened had confidence that the results were reliable and valid. Without this, it is unlikely that the screening tool would be used as it would not serve the purpose clinicians want it to.

> *'There's a group of people out there who are desperate for a diagnosis and those that support them are desperate for diagnosis. Does this become another thing they latch onto?' (C7)*
>
> *'We get families who are very sure that someone has autism and they want them to be assessed so if the screening measure came back that I don't know, they didn't seem to have autism therefore a formal assessment wouldn't be indicated, what would we then do because I suppose it is a screening measure isn't it so arguably it's not a full diagnostic' (C13)*
>
> *'The pressure is generally from people who want to know yes or no for sure do I or does my relative have autism, and so this probably isn't going to cut it with them anyway … they reject whatever score they get on the AQ-10 because they say how sure are you and I say well not very because it isn't a very good tool, and they go well do the full thing then' (C29)*

Despite identifying these potential barriers to implementing a new screening tool, when presented with a general overview of the proposed behavioural screening tool, the majority of participants were positive about the idea. In particular, that if the tool was to be successful, it would make screening more accessible to people with learning disability than it currently is.

> *'My first impression is I think it's an excellent way of observing a response because when you are just sat with a bit of paper and you're going through questions you might often miss the crucial bits' (C32)*
>
> *'I think … it will be a useful tool if it does work because I think that people are desperate out there to be honest for anything support wise for this kind of thing' (T2)*
>
> *'I can kinda see the merits in it, particularly for people who would struggle with something like the AQ cos, it's really difficult cos I always think like someone else observing you doesn't necessarily know what's going on in your head, and doesn't necessarily know what's happening for you … somebody who's got an intellectual disability and they can't kind of fill out the AQ themselves' (A18)*

Participants also saw other potential benefits of the tool, including making services more efficient, increasing knowledge about autism, and speeding up the diagnostic process.

> *'I know that the diagnostic services are all backed up so I guess from their point of view so it would maybe filter out people who … people don't meet the criteria but people think they do' (C27)*
>
> *'I think the benefits of a screening tool would be you would have a base level of knowledge which might be higher because more people might be involved in the screening tool' (C30)*

'I think one of the advantages would be in terms of efficiency I suppose thinking about it in that way. In terms of if we can screen, then we may not have to do a full assessment for somebody who isn't on the autism spectrum so I guess I think that would be useful, and likewise having that rationale to then go and do full assessment' (C13)

Teachers viewed it as something that would be of direct benefit to them, give weight to their opinion and allow children to be seen more quickly by health services.

'We're going to get something from doing it, it is actually going to support us, [so] that we have some idea of where to go next if we need the paediatrician, we need some, you know' (T4)

'I suppose as you mention a tool and so on and so forth where we could use that as part of the evidence ourselves then possibly it might be useful for us to have do an assessment of some kind which is recognised by GPs to you know accelerate that process or to pick it up and go actually straight away it's given me an idea that maybes not' (T2)

Likewise, parents and autistic people felt that such a tool would help promote earlier identification and in turn, support them.

'Anything that helps with the whole process of the diagnosis to me is a good thing' (P23)

'The sooner somebody is diagnosed the better because if, if you're sort of screening, you're putting people into the right direction … the sooner someone gets help and the guidance they need to be able to live a full life then the better' (AP12)

This theme addressed the general barriers which are anticipated when implementing a new screening tool. 'As user-friendly as possible' concerned the usability for both the person being screened and the person conducting the screening. Interviewees stressed that good psychometrics were essential to the tool being used, followed by making it as accessible as possible. Participants were positive toward the idea of video being used, although it was noted that there was a balance between a tool being accessible and being perceived as patronising. 'If it comes with a cost' looked at the more practical barriers that may stand in the way of a new tool being used. This predominantly focussed on making sure that the tool was low cost or at least cost-effective, and that training gave confidence to users of the tool. Lastly, the subtheme, 'Wedded to their own procedures and protocols,' highlighted the psychological barriers that may prevent the tool from being widely used. The dominant point here was that new tools would have to address the resistance to change within the existing systems and pathways. It appeared that the key to

getting a tool used was to demonstrate its effectiveness in making pathways and services more efficient, and how it can accurately identify people who should be referred for an autism diagnosis.

4.4 Discussion

When taken together, the results provided valuable insights into the issues to be considered when developing and introducing a new screening tool into existing systems. In general, participants expressed concerns that the current process is too lengthy and emphasised the need for timely diagnosis. This is a long standing issue, with early research by Midence and O'Neill (1999) finding that parents felt diagnosis took too long, while more recently, Crane et al. (2016) found that it takes on average three and a half years from first raising a concern to confirming a diagnosis of autism. Screening may play an important part in reducing this wait (Glascoe, 2005; McKenzie et al., 2015). Participants expressed that in order to do so, the screening tool must be quick to use, effective, user-friendly, and of practical use, which is consistent with the results of research into other clinical measures (Richardson et al., 2017). The brevity of the screening tool is very important in the context of long waiting lists, time taken to reach diagnosis (e.g. Crane et al., 2016), NHS staff shortages, and the limited time they have (Appleby & Robertson, 2016; ICM, 2016), particularly for further training and professional development (Royal College of Nursing, 2016). Any new screening tool should aim to minimise the demands on staff, both when using it and when undertaking any associated training.

The need for the tool and associated training to be cost-effective was also highlighted. NICE provide guidance about ensuring that new medical technologies are indeed cost-effective, measured against the criteria of Quality Adjusted Years of Life. There is no set threshold, but the lower the amount per year, the more desirable, as this would represent a more efficient use of NHS resources (NICE & NHS England, 2016). When designing new screening tools, this should be kept in mind.

The need for clarity about the role and remit of those involved in the diagnostic process was also highlighted. Research into inter-professional teams has emphasised the importance of role clarity to ensure that the skills and perspective of different team members can be used, to meet the diverse needs of the individual being assessed and supported (Bittner, 2018; Mickan, 2005). In order for the screening

tool to be used in practice, potential users must feel that its use is within their remit and have an awareness of what to do with the results, if the person is indicated as likely to be autistic.

There is also a need to weigh up the potential benefits of screening and subsequent diagnosis for the person before undertaking it. The participants provided a somewhat mixed picture of post-diagnostic support; while many felt that the support received post-diagnosis was beneficial, others thought that more was possible. This is also consistent with previous research. Ruiz Calzada, Pistrang, and Mandy (2012) found that a diagnosis could bring understanding and assist people to obtain the practical support they require, although the diagnosis itself did not increase the self-understanding of the people who were diagnosed as autistic, nor did it communicate exactly what kinds of support were required. Further, due to the heterogenous nature of autistic people, the usefulness of the term 'autism' alone is limited. Mockett, Khan, and Theodosiou (2011) found this mixed picture too. In their study of thirty-five families, they found that, while parents were generally satisfied with the assessment process, there was room for improvement. This included providing information about the diagnosis and how best to manage specific autistic behaviours on an ongoing basis, as well as longer and more informative follow-up appointments where they could be signposted toward post-diagnostic workshops.

The higher dissatisfaction with post-diagnostic support, compared to other aspects of the diagnostic pathway, is not uncommon. A key finding by Rogers, Goddard, Hill, Henry, and Crane (2016), who surveyed over one hundred professionals working within the diagnostic pathway for autism, was that less than half of those surveyed met NICE guidelines for post-diagnostic support, or offered a follow-up session within six weeks.

This body of research confirms the need, highlighted by participants in the current study, for there to be benefits to being diagnosed, otherwise it is a fruitless endeavour. Ip, Zwaigenbaum, and Brian (2019) provide guidelines for how autistic people can be supported post-diagnosis, including referring people for parent-mediated interventions, social skills training, and therapy to address specific issues such as co-occurring anxiety. Following such guidelines can help increase the likelihood of screening being beneficial by ensuring people receive appropriate post-diagnostic support.

Lastly, the issue of organisational barriers and resistance to change were discussed. Resistance to change in the NHS is far from a novel finding; many researchers have investigated why such a culture of resistance exists. Metcalfe et al. (2001) surveyed nearly six hundred therapists about the attitudes and barriers toward implementing evidence-based practice. Their findings show that, while the majority felt research was important for developing practice, institutional barriers prevented practice from changing due to a number of factors. These factors included: not having enough time to stay up to date with research; not fully understanding the articles which are published; that findings from research are often conflicting with other findings; not having the facilities to provide the best practice according to literature; being isolated from other knowledgeable colleagues; doctors not co-operating with evidence-based change. In a later qualitative study, Som (2005) interviewed doctors about policy initiatives and clinical governance, finding that there are many reasons why doctors may be resistant to change. Notably, some doctors felt evidence that change will have a positive impact is lacking. It was felt that clinical governance was a kind of quality assurance mechanism that does not appreciate the complexity of clinical care, rather than an agent to drive positive change, and that specific resources were not allocated to make effective changes.

In other research, Massey and Williams (2006) investigated the implementation of tools designed to support change. They reported that many staff members initially had quite extreme and emotional reactions to change but were subsequently able to benefit from it. While it is out of the scope of this thesis to explore implementation of the screening tool in detail, this chapter highlighted some of the factors that need to be considered during the development of the screening tool, to maximise the chances of it being used in practice.

Limitations

The data were collected from thirty-two individuals, all with unique experiences of the autism diagnostic pathway. Despite this heterogeneity, the themes and associated subthemes reflect consensus about key issues. The participants were from the North East of England. This means, in common with other qualitative research, that the results cannot be assumed to be generalisable to other areas or systems. Nevertheless, the identified themes were consistent with the results of previous research.

Conclusion

Overall, this study found that people frequently opt to use their gut feeling over screening tools. In respect of developing a new screening tool, the study highlighted the need for such a tool to have good psychometric properties, be clinically useful, and user-friendly. Moreover, it should be cost-effective, and minimize demands on users and those being screened, otherwise implementing it will be a challenge. Finally, even if the tool has all of these qualities, there may still be resistance to change to contend with. This study acts as a stepping-stone to the development of a new screening tool. The findings can be applied to ensure that the tool is developed with both the users and those people being screened in mind.

5.0 Chapter 5: Piloting the idea of a behavioural screening tool.

ESM: https://osf.io/e3728/ Data and outputs are broken down by study in the folder titled 'Chapter 5'

5.1 Introduction

In Chapter One, the advantages of screening were outlined, one of which was facilitating the referral of those who would benefit from an assessment for autism (Allison et al., 2012). This would help ensure that people are diagnosed appropriately and that support can be offered where it is needed. In addition, screening is beneficial for the purpose of research, where full assessment for each participant is too costly, both in terms of time and finances (McKenzie & Murray, 2015). This is the crux of why screening is needed, to ensure people are appropriately supported, and that the support provided is based on empirical research.

Chapters Two and Three highlighted the limited psychometric properties of existing instruments and the accessible adaptions of the AQ, in relation to people with learning disability, meaning that none could be recommended for use. Chapter Four showed that while screening tools are available, people opted to use their gut instinct to identify those who may be autistic. Often people's understanding of autism was incomplete, meaning this method of identification was likely to be inaccurate. More importantly, the chapter provided guidance about which properties a new screening tool should have: cost-effective, minimise demands on the person being screened and the person doing the screening, strong psychometric properties, and be of use to clinicians. This leads to the present chapter, which outlines the first steps in developing a new screening tool based upon the idea of behavioural responses.

As discussed in Chapter Three, most screening tools which exist for those with learning disability rely on the input of third-party informants. This places limitations on the extent to which the person can play an active role in their own assessment, and goes against the ideal of person-centred care (Wigham et al., 2008). The MUSAD (Bergmann et al., 2015) however, is an exception. This uses observable responses to musical tasks to screen for autism. The current study further explores the idea of using behavioural responses to stimuli as an autism screen. Here, a mixture of both observable and reported responses to a

specific set of stimuli are used; the idea being that these responses may form the basis of a score on a screening tool.

The overall aim of this chapter was to develop proof of concept, for the idea of a behavioural screening tool. The associated aims of the four studies presented in this chapter are to: identify potential stimuli informed by existing literature, that are likely to provoke different responses depending on the autistic traits or diagnostic status of the participant; provide an insight into how a screening tool may work in practice; and highlight some of the issues which should be anticipated when conducting the main study, presented in Chapter Six. The four studies which are presented investigated different aspects of a potential behavioural screening tool. The literature that informed the creation of the stimuli used in the studies is outlined below.

5.1.1 **Empathy: What is it?**

It has been argued that empathy is at the heart of human behaviour (Smith, 2009). The concept was first identified in a German paper translated to English a little over one hundred years ago (Preston & de Waal, 2001). Despite its relatively longstanding usage, the concept did not receive significant attention from researchers until the mid-twentieth century (Cottrell & Dymond, 1949). There is little consensus about the best way to define empathy. Stotland (1969) defines it as a vicarious emotional response to another person's emotion; in essence, the observer shares the feelings of the person being observed. In the same year, Hogan (1969, p. 307) defines empathy as 'the [accurate] intellectual or imaginative apprehension of another's condition or state of mind without actually experiencing that person's feelings.' While the difference may appear minor, the implication is profound: Stotland argued that empathy is the act of sharing and feeling another person's mental state, while Hogan argued that it is only an awareness and understanding of that person's state of mind. This debate has continued in the subsequent years. What is perhaps clearer is the distinction between empathy and sympathy. The latter, can be thought of as having feelings for another person or the situation which they are in, whereas empathy is feeling as though you were in that situation or simply having an awareness of how someone in a situation may feel (Hein & Singer, 2008). Only empathy is of interest to this chapter.

The discrepancy between research about what exactly empathy is, has led to two differing approaches being taken: the affective approach in line with Stotland, and the cognitive approach, in line with Hogan. In terms of affective empathy, Baron-Cohen and Wheelwright (2004) state that there are multiple types discussed in previous literature. Firstly, the observer's emotion matching that of the observed. Secondly, the observer having an appropriate feeling toward the person being observed. Thirdly, any feeling deemed appropriate from the observer toward the observed person's emotional state; for example, experiencing pleasure at another person's distress is inappropriate. Lastly, empathy includes feelings of compassion toward a person's distress.

Conversely, cognitive empathy is based around the understanding of another person's feelings and point of view. In this case, Baron-Cohen and Wheelwright (2004), state that the terms 'Theory of Mind' (see: Baron-Cohen, Leslie, & Frith, 1985) and 'Mindreading' (see: Baron-Cohen, 1995) refer to the same set of skills. These are skills that allow predictions to be made about a person's emotional state and for an observer to empathise with the person being observed. These skills do not, however, necessitate that the observers feel the observed emotion. These two similar, yet differing conceptualisations of empathy, form the basis of how it is understood within current literature.

The debate about empathy continues today, with some researchers adopting a purist cognitive or affective approach, whereas others view it as a mixture of both (see: Cuff, Brown, Taylor, & Howat, 2016). While the two elements may appear separable, Cuff argues that they are not. Notably, Strayer (1990) suggested that cognitive empathy provides a process and pathway for the creation of affective empathy. Complementary to this, work by Lamm, Batson, and Decety (2007) found that when cognitive elements are manipulated, the affective components are affected. While Cuff and colleagues (2016) outline this debate in far greater detail, they conclude that empathy is indeed a composite of both cognitive and affective components. Henceforth empathy here is defined as:

> *'An emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with recognition that the source of the emotion is not one's own.' (Cuff 2016 p. 150)*

Researchers have attempted to translate empathy into a measurable and quantifiable construct. Much like the definition, debate exists about whether empathy is measured as purely cognitive, purely affective, or as a mixture. Accordingly, multiple measures have been developed which mirror these viewpoints. This makes the measures difficult to compare, as they measure various facets of empathy. Hogan's (1969) definition of empathy is in line with cognitive empathy, and therefore the Hogan Empathy Scale focusses on this, to the absence of any affective components. On the other hand, Baron-Cohen and Wheelwright's (2004) definition focusses on the emotional response of a person, and as such their Empathy-Quotient (EQ) taps into affective components. More recently, tools that amalgamate the two constructs have been developed. One example of this is by Reniers, Corcoran, Drake, Shryane, and Völlm (2011), who developed the Questionnaire of Cognitive and Affective Empathy. This allows the measurement of each construct through subscales and provides an overall score, including both cognitive and affective components. The lack of consensus needs to be considered in relation to research findings.

In all, empathy is a diverse construct which may underpin many aspects of human behaviour but, because of its broadness, attention to what kind of empathy is being discussed and researched in specific circumstances is needed. Nevertheless, empathy has been related to various aspects of behaviour and psychology.

5.1.2 **Empathy: Its relationship with autism**

Research by Baron-Cohen (1995) into empathy in autistic people discusses autism in terms of mind-reading and Theory of Mind, which were later clarified by the author as analogous to empathy (Baron-Cohen & Wheelwright, 2004). Baron-Cohen's 1995 argument presents the idea that non-autistic people unconsciously and frequently mind-read, which allows them to have insight about another person's thoughts, feelings, and intentions. Crucially, he argues that autistic people do not do this.

Various strands of evidence provide some support for this claim. Baron-Cohen et al. (1985), used the 'Sally and Anne' puppet task to show that children with autism do not attribute differential knowledge to different people. In turn, this suggests that those children do not draw upon Theory of Mind to make decisions. Autistic people were also found to consistently score lower in terms of empathy, as measured by the EQ, compared to non-autistic people (Baron-Cohen & Wheelwright, 2004). In a general review of

autism literature, Theory of Mind and the notion of 'mindblindness' (i.e. poor Theory of Mind) are used as explanations for many of the social deficits seen in autism (Frith & Happé, 2005). Similarly, a review discusses the empathy impairments in autism, concluding that many observed social deficits are attributable to disrupted empathetic ability (Chen, 2012). With mounting evidence, discussion of empathy disruptions in autism is common (see: Bird & Viding, 2014; Lawson, Baron-Cohen, & Wheelwright, 2004).

Other researchers, such as Tager-Flusberg (2007), acknowledge that Theory of Mind impairments and reduced empathy are common in autistic people, but highlight studies which show that, at times, neither autistic adults nor children show any Theory of Mind deficits. Additionally, Fletcher-Watson and Happé (2019) make the point that Theory of Mind merely develops later in autistic people, rather than not at all.

Notably, empathy differences in autism seem to predominantly exist within the cognitive domain. In a meta-analysis, Song, Nie, Shi, Zhao, and Yang (2019) reviewed the area and broke empathy down into three subtypes: cognitive, empathic accuracy, and empathic concern, with the latter referencing specific emotional responses to suffering. Each subtype was then divided into a trait level, meaning a universal global trait, or state level, meaning within a specific context. This leads to six subdomains: cognitive trait; cognitive state; empathic accuracy trait; empathic accuracy state; empathic concern trait; empathic concern state. When autistic people were compared to non-autistic controls, they were indeed found to be poorer on four of these subdomains: cognitive trait, cognitive state, empathic concern trait, and empathic concern state. Meanwhile, autistic people were found to be comparable to controls on empathic accuracy state and in fact, superior on empathic accuracy trait. This review indicates that autistic people potentially have as good, or better, affective empathy than non-autistic people.

The overall body of evidence shows that, in some circumstances, autistic people seem to differ in terms of empathy when compared to non-autistic peers. In terms of a behavioural screening tool this is crucial as, if there is an underlying difference in empathy, it is possible that this can be observed by down-stream behaviours resulting from these empathy differences.

5.1.3 **Mimicry and emotional contagion's relationship with empathy**

Analogous to empathy is mimicry and subsequent emotional contagion. Humans tend to express congruent emotions when viewing emotional faces (Dimberg, 1982), and mimicry is the apparent expression of that emotion. According to Hess and Blairy (2001), mimicry is the copying of a viewed expression without necessarily internalising the observed emotion. In a review of facial mimicry research, Hatfield, Bensman, Thornton and Rapson (2014) highlight that this is ubiquitous throughout the world, although according to Markus and Kitayama (1991), culture may influence exactly how it is expressed. Markus and Kitayama argue that mimicry is more common in cultures that value connection, such as Asian cultures, whereas those that have a greater preoccupation with the self, such as American, are less prone to mimicry. Regardless of how exactly mimicry is displayed, Hatfield et al. (2014) point out that many studies find no cross-cultural differences in mimicry and that it is relatively stable worldwide.

Hatfield et al.'s (2014) review highlights that, in empirical research, an observer's face will change in a manner congruent to that being viewed, irrespective of whether the stimulus is a static image or a dynamic video. This is consistent even when people are not prompted to react to a face. Moody, McIntosh, Mann, and Weisser (2007) demonstrated this idea when they asked participants to watch a series of movie clips without further instructions and recorded their muscle movements using electromyography (EMG). The EMG showed that the participants subtly reacted in a manner in line with the emotion being viewed. A further EMG study found similar results, showing that people will spontaneously react when exposed to emotive stimuli (Dimberg, Thunberg, & Elmehed, 2000). In an earlier review, Hatfield et al. (1993) cite evidence that mimicry is innate due to it occurring so soon after birth. For example, Jordan and Thomas (2017) show that mimicry of positive affect occurs within the first twenty-four hours after birth. Taken together, there is a body of evidence that facial mimicry happens naturally and is an innate response that is expressed at a very early developmental stage.

Mimicry provides a potential explanation as to how empathy occurs. As at least a part of empathy is feeling emotions which someone else is showing (Cuff et al., 2016), mimicry can assist this process. Hatfield, Rapson, and Yen-Chi (2009) propose that there are a number of steps to the empathy process:

first, ‘mimicry’, where the observer copies an expression; second, ‘feedback,’ where they feel the mimicked expression; finally, ‘contagion,’ where they adopt the affective state. This theory is in line with the Perception Action model of empathy (Preston & de Waal, 2001). Hatfield et al.’s (2009) proposed mechanism allows someone to take, understand, and empathise with another person’s perspective. This leads to emotional contagion, as the individual will have felt the emotion of another person in that third step. Emotional contagion is a pertinent idea here, as in part, it can be thought of as an outcome of the process. Hatfield et al. (1993) defined emotional contagion as the tendency for an observer to mimic and synchronise expressions, vocalisations, postures, and movements with someone, then consequently, converge emotionally with the observed expressed state. Moreover, Hatfield and colleagues (2014) argue that contagion forms an important part of interactions as it helps people understand the thoughts and feelings of each other. In a sense, this means that mimicry can lead to emotional contagion and in turn, provide a mechanism for empathy to occur. This makes both mimicry and the potential contagion that follows, important aspects of the empathising process.

In empirical research, it has repeatedly been shown that emotional contagion can be produced under experimental conditions. Barsade (2002) randomly assigned participants to groups and subsequently to experimental conditions, where a confederate blind to the hypotheses of the study was instructed to perform affective actions. There were four different conditions that people could be assigned, either Cheerful Enthusiasm, Serene Warmth, Hostile Irritability, or Depressed Sluggishness; with the two former being pleasant conditions and the two latter, unpleasant conditions. Barsade assessed mood by both video coding, and self-report questionnaires, finding that the participant’s mood was indeed influenced by the confederates, but less so in the negative conditions. Hennig-Thurau, Groth, Paul, and Gremler (2006) had participants take part in staged customer service interactions. The person providing the ‘service’ to participants in this interaction was a confederate, who was a trained actor, and instructed to either smile a great deal or smile a little, and to either act convincingly or unconvincingly. Using pre- and post-encounter questionnaires, Hennig-Thurau and colleagues found that people were influenced by the affect of the person that they interacted with and would feel more positive after the encounter, but

this was not mediated by how much smiling the person did. Instead, it was driven by how convincing the person's acting was, with more convincing acting leading to higher reported affect.

Emotions have also been found to be consistently contagious outside of the laboratory. Similar to Hennig-Thurau's study, Pugh (2001) investigated emotional contagion in retail experiences, in real-world scenarios. Across regional branches of a bank, customers and staff were observed and surveyed using standardised questionnaires. After linking the staff to customers, the findings indicated that the emotions displayed by staff consequently affected the customer's emotions. In a different setting, Kramer, Guillory, and Hancock (2014) investigated emotional contagion on social media by manipulating emotional expressions on people's Facebook 'news feed,' and then assessing the person's posting habits on the site. Kramer and colleagues found that those exposed to more positive posts posted more positive content and comparably, those exposed to more negative posts posted more negatively. This shows that emotional contagion is complex and can occur in numerous ways. Importantly, this also means that emotional contagion is not limited to the emotional expressions people see, as has been shown in other research. Two reviews by Hatfield and colleagues (Hatfield, 1994; Hatfield et al., 2014) provide a comprehensive overview, both concluding that there is clear evidence that exposure to emotional stimuli leads to people 'catching' the observed emotion.

One potential mechanism for this emotional contagion is mimicry, and many have investigated this link. Arguably, mimicry is a tool that can be employed to increase prosocial behaviour, to make someone act in a more helpful manner, and to be kinder (Chartrand, Maddux, & Lakin, 2005; Van Baaren, Holland, Kawakami, & Van Knippenberg, 2004), as well as increase empathy between people (Hasler, Hirschberger, Shani-Sherman, & Friedman, 2014). One argument, similar to that proposed previously, is that mimicry can help someone understand another person's emotional state via a feedback mechanism, that allows a person to understand (cognitive empathy), and feel (affective empathy) what another person is feeling (Neal & Chartrand, 2011; Niedenthal, 2007). The converse argument is that mimicry occurs due to the person first understanding and feeling that emotional state, then displaying it (Hess & Fischer, 2013). While the direction of the relationship between empathy and mimicry is unclear, what has been shown throughout research is that the two are certainly interconnected.

Generally, it is thought that those with high levels of affective or emotional empathy show higher levels of mimicry (Seibt, Mühlberger, Likowski, & Weyers, 2015). Sonnby–Borgström (2002) found that those with higher levels of empathy exposed to images of faces performed more mimicking behaviour, as well as showing a higher correspondence between facial expressions and reported feelings. Rymarczyk, Żurawski, Jankowiak-Siuda, and Szatkowska (2016) used video clips of emotional expressions and found that participants who scored higher on traits of empathy showed greater reactivity and mimicry to stimuli. Likowski, Mühlberger, Seibt, Pauli, and Weyers (2011) made very similar findings which showed that cognitive empathy could predict mimicry, but could also be influenced by the situation the person is in. For instance, if the person is in a competitive situation, such as a game of dice, the person may not show mimicry but instead perform another emotion in response to their opponent's emotion. In all, while this mechanism, and the reasons behind it, are not well understood, it is clear that a relationship between empathy and mimicry exists.

Although this research indicates that emotional contagion is common, it should be noted that while mimicry and emotional contagion are often congruent, it is not necessary that they are. Hess and Blairy (2001) used a set of video stimuli which they presented to participants, and measured both their mimicry and contagion of the viewed emotions. Hess and Blairy found evidence for both mimicry and emotional contagion in the participants; however, they did not find that the two were systematically related in any way, meaning that one could be present without the other. In turn, no evidence was found for mimicry facilitating contagion, nor contagion facilitating mimicry. Hatfield and colleagues (2014) argue that the two are likely linked but this has yet to be definitively shown throughout the literature, citing evidence from studies showing that people are capable of producing either emotional contagion or mimicry, but not always both.

This research shows that empathy differences may indeed lead to differences in mimicry, and due to subsequent emotional contagion, there may be variation in the emotions that people feel. In this sense, both observation of action using video, and asking questions around how a person feels, could reveal information about their empathic tendencies. Though, it is important that the displayed emotions are convincing as the believability can heavily influence the downstream contagion.

5.1.4 **Mimicry and autism**

Research shows that autistic people have disrupted empathic abilities, and because of the link between empathy and mimicry, the relationship between autism and mimicry has also been investigated. McIntosh, Reichmann-Decker, Winkielman, and Wilbarger (2006) report that early studies comparing children with and without autism often found no differences in terms of mimicry. However, these early studies were not focussed on facial and emotional mimicry but were instead action-orientated, whereby a person observed an action and was asked to imitate it (e.g. pointing; see McIntosh et al., 2006; Rodgers, 1999). Some studies do show differences in terms of action imitation and mimicry. Jiménez, Lorda, and Méndez (2014) employed an imitation task, presented via video, where children were asked to copy the movement and location of coloured bars. When children were asked to do this, they found that both those with and without autism showed an ability to replicate the end goal, albeit they achieved this in a different way. The children with autism struggled to replicate the exact movements and steps shown but were able to achieve the final end goal in a more efficient manner than the children without autism. Similar research has extended to adults, using virtual avatars presented on a computer. Forbes, Pan, and Antonia (2016) found that both autistic and non-autistic adults were able to mimic movements viewed on a screen, but autistic people tended to mimic to a lesser extent. That said, imitation tasks such as these are different to the emotional mimicry previously discussed.

Many researchers have investigated facial emotional mimicry. Early work by Hertzig, Snow, and Sherman (1989) used photographs of faces displaying affective states, and also expressions performed by the researchers. Hertzig and colleagues found that when the children were asked to reproduce affective states when looking at pictures, both children with autism and learning disability performed worse than the control group. Yet, when asked to imitate the affective state shown by the researcher, the children with learning disability performed comparably to the control group, whilst the children with autism performed significantly poorer. Further research by Loveland et al. (1994) asked children with either autism or Down syndrome to imitate expressed emotions acted out by researchers, and to show emotions based on verbal instructions. They found that the two groups performed comparably on the imitation task, but that the children with autism produced less recognisable emotions when being asked to create

them from verbal instructions. While the two performed comparably on the imitation task, it should be kept in mind that the participants were cued to produce emotions, and that the study itself lacked a typically developing control group to compare both groups to. Later, Scambler, Hepburn, Rutherford, Wehner, and Rogers (2007) compared a group of children with autism, a group with developmental delay, and a typically developing control group on reactions to facial expressions presented by a researcher. Scambler and colleagues video-recorded the child's response and only analysed the expressions that the children actually attended to. They found that the children with autism reacted with congruent emotions about half as often as the other two groups. As only the trials where children attended the stimuli were included, the authors argue that the differences cannot be due to differences in attention. McIntosh et al. (2006) showed autistic participants images depicting a range of facial expressions, initially without any prompts, and subsequently while being asked to copy the expression. They found that autistic people did not mimic the expressions when unprompted but did when asked to copy them. This indicates that while autistic people are capable of mimicry, it is not automatic. Similarly, using EEG, Oberman, Winkielman, and Ramachandran (2009) found that autistic participants, while performing comparably to non-autistic people in terms of voluntary and conscious mimicry, were slower in terms of spontaneous mimicry. Other studies make similar findings with autistic adults. Yoshimura, Sato, Uono, and Toichi (2015) showed participants videos which were used in previous mimicry EEG based research (Sato & Yoshikawa, 2007), video-recorded the person's face, and subsequently coded their reactions. Their findings show that it is possible to visually identify spontaneous mimicry and observe that autistic adults display lower levels of it. Crucially, in a review of the literature, Vanvuchelen, Roeyers, and de Weerdt (2011) argue that the differences seen in mimicry and imitation are to some extent influenced by developmental delay, but intellectual ability alone does not wholly explain the difference, making these differences unique to autism and not attributable to other conditions.

While this research shows autistic and non-autistic people may differ in terms of mimicry, little direct research has been conducted on whether emotional contagion from facial expressions is present in autistic people. Some related research can provide insight. Using fMRI, Hadjikhani et al. (2009) presented body language expressions of fear and neutral poses to participants. They found evidence that,

compared to non-autistic people, autistic people show a reduced difference in reaction to either neutral or fear stimuli. As such, Hadjikhani and colleagues argued that this represented evidence of reduced contagion of fear in autistic people. In a later fMRI study, Hadjikhani et al. (2014) assessed reaction to pain in a similar manner. Here, videos of neutral and pained faces were shown to both autistic and control participants. The findings revealed activation in both groups' pain matrix and no statistical difference between the two groups, and as a result, the authors argued that contagion of pain remains intact in autistic people. While directly relevant research is rare, in related research, when differences in mimicry and empathy are taken into account, there is the suggestion that some differences in emotional contagion will be present in autistic people, in comparison to non-autistic people. This means that when autistic and non-autistic people view the same sets of stimuli, differences in what they report feeling might be expected. However, past literature does not provide any real certainty on that aspect.

Overall, the research discussed here indicates that autistic people may present differences in the way that they mimic when responding to emotional stimuli. It appears likely that mimicry is reduced, but there is less certainty about whether comparable or less emotional contagion will take place. That said, when the differences with mimicry are viewed alongside research on empathy, it is probable that a reduction in emotional contagion should be expected.

5.1.5 **Mimicry of laughter**

A related, but more narrow, incidence of mimicry and contagion relates to laughter. Provine (1996) argues that any act can synchronise group behaviour, and highlights laughter as a specific social signal that people can synchronise around. Laughing is inherently social. Humans are estimated to be approximately thirty times more likely to laugh in the presence of others compared to when alone, and the speaker is more likely to laugh than the listener, indicating that it is more than a simple reaction to humour (Provine, 2004; Provine & Emmorey, 2006). A variety of research confirms the contagious nature of laughter. One study shows that listening to a box producing laughter is enough to elicit a laughter response in people (Provine, 1992). Another presents evidence that infants less than twenty-four hours old respond to laughter, with laughter (Jordan & Thomas, 2017). In real-world situations, such as theatres, audience size predicts the level of laughter exhibited by individuals, with the presence of more

people driving the contagious nature of laughter (Butcher & Whissell, 1984; Levy & Fenley, 1979). It should be noted that while laughter is contagious, repeated exposure to stimuli of laughter has a diminishing effect; people may instead find it irritating, making them less prone to respond with laughter (Provine, 1992).

Laughter is considered by some as an emotion (e.g. Scott, Lavan, Chen, & McGettigan, 2014), and Jung (2003) argues that for laughter to be triggered, the subject must understand both the other person's inner state and the inner cause of the laughter. Therefore, a potential relationship between empathy and the contagious laughter response might be expected. While much of the research in this area is considered more generally with emotional contagion and mimicry (e.g. Hatfield et al., 2014), specific research in this area does exist. Neves, Cordeiro, Scott, Castro, and Lima (2018) found that being able to assess the authenticity of laughter was related to scores on contagion and empathy scales, and responsively laughing at the stimuli. As with other forms of mimicry, laughter and autism have been researched. Hudenko, Stone, and Bachorowski (2009) found that children with autism perform comparably with their peers in terms of producing laughter that reflects internal states, for example, responding to what they deem humorous. These same children, however, exhibited a diminished use of laughter as a social signal.

As with other forms of mimicry, contagious laughter appears to be reduced in autistic people. Reddy, Williams, and Vaughan (2002) investigated a group of children with autism and a developmentally matched group of children with Down syndrome. Parents interacted with their children during video-recorded naturalistic interactions of free play and play-with-toys sessions at home, across multiple visits. These parents were briefed by researchers about what behaviours they should look out for, including attempting to join in with other people's laughter, repeating acts they had previously seen as eliciting humour, and playful teasing. Parents were also provided with Dictaphones to record anything of interest to the study that occurred between visits. According to parents, laughter in response to funny faces was reduced in children with autism. When viewing the video recordings of the play sessions, the children with autism were less likely to both attend to and react to other people laughing. This illustrates decreased attention to and mimicry of laughter, compared to the children with Down syndrome. Helt and Fein (2016) found that children with autism were less affected by the presence of other people, by laugh

tracks imposed on a cartoon and by using a pencil to force smile muscles into activation, as means for evoking laughter, compared to mental age matched controls. The first two indicate that other people and the sound of laughter are less influential for children with autism, as compared to children without. The latter point, when considered in the context of a feedback loop, such as that proposed by Hatfield et al. (2009), indicates that in children with autism there may be an issue with the feedback step. Taken together, this research supports the expectation that, as laughter can be considered an emotion, reduced mimicry of laughter in autistic people should occur.

As shown here, laughter is not merely an action people exhibit in response to humour, it is also a social cue. Additionally, this indicates that in the same way as mimicry occurs with expressions of facial emotions, mimicry also occurs with laughter. Just as mimicry appears to be generally different in autistic people, evidence indicates that there are likely to be differences in contagious laughter too. This is important, as laughter as a stimulus may prove a helpful addition to a behavioural screening tool.

5.1.6 **Mimicry of yawning**

Related to mimicry and emotional contagion is the contagion of specific actions, such as yawning. Provine (2005, p. 1) argues that the 'yawn is primal, unstoppable and contagious.' Firstly, it can be considered primal as the contagious yawn is present in primates (e.g. Deputte, 1994; Palagi, Leone, Mancini, & Ferrari, 2009), as well as new-borns and foetuses (Sepulveda & Mangiamarchi, 1995). Secondly, in terms of it being unstoppable, Provine (1986) theorises that yawning should be thought of as a fixed action pattern. In this case, the yawn is an invariant behavioural sequence that once initiated, must run until completion and is unable to be stopped. Lastly, yawns are contagious, as once someone thinks about, reads about, or sees another person yawn, they themselves will likely yawn (Provine, 2005).

There are two separate forms of yawn which develop at different times. The earliest is spontaneous yawning, found in new-borns and foetuses (Provine, 2005). Explanations for this type of yawn vary, with some researchers stating that it increases arousal and motivation (Walusinski, 2006), and others that it regulates brain temperature (Gallup & Gallup, 2008). However, it is the second form of yawning which is of interest here.

The second form is contagious yawning, which is well documented throughout the literature. One classic study by Provine (1986) shows that contagious yawning can occur from both seeing yawns and reading about them. Contagious yawning has been reliably produced in various experiments involving both humans and animals (for a review see: Guggisberg, Mathis, Schnider, & Hess, 2010). This form of yawning appears to develop later, generally around ages four to five (Anderson & Meno, 2003; Giganti & Esposito Ziello, 2009).

Studies have linked various factors to contagious yawning. Platek, Critton, Myers, Gallup Jr., and Gallup (2003) found yawning was negatively associated with schizotypal personality traits and positively associated with participants reacting to their own face and performance on a theory of mind task (i.e. cognitive empathy). This link to empathy is crucial and relates to the potential social basis of contagious yawning. It has also been argued that contagious yawning provides a means of communication (Deputte, 1994), and an empathic connection (Paukner, Suomi, Visalberghi, & Ferrari, 2009). Due to this, contagious yawning is also known as social yawning. It has been speculated that this occurs by mimicking the start of a facial expression, in turn triggering the aforementioned fixed action pattern and leading to the completion of a yawn (Helt, Eigsti, Snyder, & Fein, 2010). This is comparable to the mimicry, feedback, contagion model of empathy outlined by Hatfield et al. (2009). Preston and de Waal (2001) make the case that yawning can be involved in the perception action system, as the yawn is observed, and the reflex is the automatic imitation of the process. In addition, the brain areas associated with empathic processing are also activated during contagious yawning (Platek, Mohamed, & Gallup, 2005; Schürmanna et al., 2005), emphasising the interconnectivity of empathy and contagious yawning. While this connection has been highlighted within the literature, other researchers have argued that the two are not equivalent, and that this area of research is not conclusive at this time. Massen and Gallup (2017), in a review of the literature, highlight that while evidence of this connection exists, it is inconsistent, with numerous studies finding no relationship. Adding to this, studies often have confounding variables, like visual attention or social inhibition, which can impact the outcome. Massen and Gallup argue that this connection needs more robust evidence to ascertain the precise nature of any relationship.

Research indicates that autistic people appear to experience differences, both in terms of empathy and mimicry, and that these differences extend into contagious yawning, even if the precise nature of the empathy-yawning relationship is still uncertain. To the researcher's knowledge, the earliest study investigating this was conducted by Senju et al. (2007) where a sample of twenty-five children with autism and twenty-five age and gender matched controls viewed yawning and control videos (a person opening their mouth). Findings revealed that the children with autism yawned less than control participants when viewing the yawn stimuli, and that the number of yawns performed by the children with autism did not change depending upon whether they viewed the yawn or control stimuli. These results were replicated in an IQ-matched subsample. Helt et al. (2010), also replicated these general findings across a series of studies where yawning was performed in person, whilst the experimenter read a story to a group of non-autistic children aged one to six years old. Results showed that children under four years old were unlikely to contagiously yawn, but beyond that no influence of age was found. In a second study which included children with and without autism, the researcher read to the children individually, yawning four times while doing so. Findings showed that children with autism were less likely to yawn compared to non-autistic controls. Giganti and Esposito Ziello (2009) performed a similar study but separated the children with autism into composite groups, which they titled 'high-functioning autistic children' and 'low-functioning autistic children,' and compared them to a non-autistic control group. Spontaneous yawning appeared consistent between all groups, but contagious yawning was increased in the non-autistic group when viewing videos of and listening to the sound of yawns. The two groups of children with autism did not differ from each other. This finding highlights the role of both visual and auditory stimuli in contagious yawning. Together, these studies indicate a clear reduction in contagious yawning between autistic people and non-autistic controls.

One explanation for the differences in contagious yawning between autistic and non-autistic groups, other than differences in empathy, relates to eye contact. Reduced eye contact in autism is well researched (e.g. Chawarska & Shic, 2009; Pelphrey et al., 2002), and Senju et al. (2009) citing Provine (1989) raises this, as a potential explanation for contagious yawning differences. The 2009 study used the same stimuli employed by Senju et al. (2007), but a fixation cross was placed where eyes would

appear, cueing participants to look at that region when the yawn stimuli appeared. This led to the autism and control groups showing equivalent contagious yawning. Further, Usui et al. (2013) used eye-tracking to confirm that when autistic participants viewed the eye region, and to a lesser extent the mouth, contagious yawning was more likely. No effect of IQ, age, or autistic traits were found. These results indicate that a potential factor in contagious yawning is the attention to specific facial features, although this cannot fully explain reduced contagious yawning by autistic participants found by Giganti and Esposito Ziello (2009) when only using auditory stimuli.

Aside from eye contact, Bartholomew and Cirulli (2014) investigated a number of variables related to contagious yawning in a non-autistic sample. Their findings showed that only age had any significant explanatory power (8%), and that empathy, emotional contagion, circadian energy rhythms, and sleepiness did not. Further, Chan and Tseng (2017) measured autistic traits and split the non-autistic sample into high and low scorers. The two groups performed comparably, and no significant differences were found. These findings imply that the consistently observed deficits in contagious yawning are reserved for those diagnosed with autism. The reasons for this remain unclear.

Overall, it appears paying attention to facial features assists contagious yawning, but that contagious yawning is not limited to visual stimuli and can occur in response to only hearing the sound of a yawn. Additionally, although empathy appears to play a role in contagious yawning, it is unlikely to provide a full explanation as to why contagious yawning occurs, as some research fails to show any relationship between the two. Finally, although some research has indicated differences in contagious yawning between autistic and non-autistic controls, research has not shown a relationship between contagious yawning and autistic traits. Autism and reduced contagious yawning appear related but the exact mechanism as to why is unknown.

In summary, yawning is more than a mere indicator of tiredness. Instead, there are social qualities to it. As with the previously discussed emotions, shown by facial expressions and contagious laughter, yawning also appears contagious. As stated, there appears to be differences between autistic and non-autistic people in terms of contagious yawning. Accordingly, yawn stimuli may provide a further useful type of stimulus for the screening tool.

5.1.7 **Cuteness**

In research with infants and caregivers, cuteness appears to be crucial. In a review by Kringelbach, Stark, Alexander, Bornstein, and Stein (2016), they ascertained that cute infants quickly capture the attention of adults, regardless of relationship, and in turn the cute infant would receive beneficial caregiving. Most research focusses on cute faces, finding that a head which is too large for the body, with large eyes, plump cheeks, high eyebrows, and a small chin are factors that makes a face cute. These are features that infant humans and many animals have. Kringelbach and colleagues (2016) argue that this cuteness can go further than prompting simple caregiving, also giving rise to empathic reactions from others as cuteness promotes prosocial behaviour. This cuteness can lead to an increase in empathy for the cute entity. Further, it can be argued that cuteness promotes the humanisation of entities and this results in a range of behaviours that care for and support that entity (Sherman & Haidt, 2011). These theories extend to animals, as many animals appear cute to humans through an evolutionary adaption, and can result in humans caring for the animals in question (Bradshaw & Paul, 2010).

As people's reaction to cute entities seem related to empathy, there is a potential for differences to exist here between those who are and are not autistic. Though overall, this area is under-researched and to the researcher's knowledge, no evidence exists that directly addresses the question of whether reactions to cuteness differs between autistic and non-autistic people.

5.1.8 **Empathy for animals and humans**

If the theory of contagious reactions being a product of empathy are true, one area to consider is that people may show empathetic reactions to both animals and humans, though there is limited research in this area. Westbury and Neumann (2008) conducted some early work to test for differences in empathy towards humans and animals when viewing videos of both in negative situations. In the sample of seventy-three participants, subjective measures of empathy and phasic skin conductance response (a biological measure of empathy) were shown to be related. Self-rated empathy towards humans was found to be higher than that felt towards animals, while the amount of empathy towards animals was found to differ depending on how human-like the animal was. For example, greater empathy was shown toward mammals compared to birds. Additionally, Franklin et al. (2013) conducted an fMRI study where

participants viewed both humans and animals suffering. Results showed that there was greater activation in areas of the brain related to empathy when viewing humans compared to animals. This shows that, although people appear to react more strongly to other people compared to animals, they still have some empathy for animals.

In related research, there is much evidence to show that autistic people are able to bond with animals. Early pilot work carried out by Sams, Fortney, and Willenbring in 2006 found that during sessions, the use of therapy animals of various sorts, such as llamas, dogs, and rabbits, with children with autism resulted in increased language and significantly greater social interactions. This indicates that animals can have beneficial effects for children with autism. O'Haire, McKenzie, Beck, and Slaughter (2013) investigated this further by video-recording play sessions between children with autism and their typically developing peers. Participants were compared in terms of play with either toys or guinea pigs. The researchers found that in the presence of guinea pigs, children with autism displayed more prosocial behaviour and were approached by non-autistic peers in more prosocial ways, compared to when they played with toys. These results indicate that, to some extent, animals may have a positive impact on social outcomes for autistic people, both in terms of being more sociable and also being approached more often by their peers. Later work by the same group of researchers measured physiological arousal in children with and without autism, who were again playing with guinea pigs (O'Haire, McKenzie, Beck, & Slaughter, 2015). They found that, compared to reading silently, reading aloud, and free play with toys, the children with autism showed lower levels of arousal when playing with the guinea pigs. According to the authors, this indicates the calming effect of the animals which provides a social buffer for the children. However, the calming effect of animals was not shared to the same extent in children without autism. Separately, in interviews with parents of children with autism, Byström and Persson (2015) found that parents felt their children benefitted from companion animals. Such findings are common amongst animal interventions for autistic people, with a review showing that these interventions often lead to improved social experiences (O'Haire, 2017). In all, this research indicates that animals have a positive and beneficial impact on autistic people, which is greater than that experienced by their non-autistic peers.

Together, this evidence shows that, largely due to empathy, people are likely to react to animals and humans in a different manner. Next, there are likely differences between autistic and non-autistic people in their reactions to animals, as animals seem to have some additional therapeutic and beneficial effects for the former. This indicates that if stimuli of both animals and humans are included in a behavioural screening tool, it is possible that a greater variation in responses will be seen.

5.1.9 **Rationale and aims**

As discussed in Chapter Three, current screening tools when used with people with learning disability are inadequate for identifying people who are likely to be autistic. These tools often rely on third-party raters to provide information about the person. For some people, this simply is not possible as they may not have others who know them well enough to respond on their behalf, as staff teams are often unstable due to the high staff turnover in many learning disability care settings (Butler et al., 2010; Robertson et al., 2005). This highlighted the need for a different type of screening tool. The literature outlined above showed that there are reliable, quantifiable differences between autistic and non-autistic people in specific and observable behavioural responses. This included reactions to faces showing emotions, contagious yawning, contagious laughter, and potentially in empathy relevant reactions such as cuteness. Further to this, the inclusion of both animal and human stimuli in the research have led to a broader range of reactions, and in turn aided the accuracy of the screening tool. In theory, these reactions could be performed by a range of people, including those with learning disability.

This study aimed to be a proof of concept and to explore whether it may be possible for an observation-based measure to be used to categorise people who are likely and unlikely to be autistic. The chapter described a series of pilot studies during which participants, who differ in respect of their levels of autistic traits (measured by the AQ), viewed a range of dynamic stimuli. These were in the form of pre-existing videos, that were thought to likely elicit responses which people can self-report. The results of the four pilot studies presented here were also used to inform the content and delivery of the final behavioural screening tool. Moreover, the studies outlined the development of a scoring system, and assessed if differences exist between third-party and self-rating. Taken together, the studies attempted to

explore whether a person's behavioural reactions to video clips could accurately estimate their score on the AQ.

5.2 Method & Results

5.2.1 General method

Video Stimuli

The stimuli were pre-existing videos sourced from internet video websites (YouTube and Vine). Dynamic stimuli were chosen based on research that indicated such stimuli has a greater influence on people's empathic reaction than static stimuli (Hatfield et al., 2014; Rymarczyk et al., 2016).

The content of the videos was selected based on the literature outlined above, meaning they were potentially suitable to evoke a range of reactions from participants and these reactions could potentially discriminate between those with different levels of autistic traits. This idea was based upon likely underlying differences in empathy and associated differences in contagion and mimicry. As such, the stimuli included videos depicting yawning, laughter, awkwardness, confusion, pain, disgust, the response to sour food, and receiving a scare. Videos aiming to elicit reactions in response to cuteness were also included. The video content included: Yawning; Laughter (e.g. a child laughing at a dandelion being blown); Awkward (e.g. having a kiss rejected); Confusion (e.g. singing in high pitched voice in a foreign language and a flower being placed on a cat's head); Sour Food (e.g. a child eating a lemon); Cute (e.g. kittens, babies); Disgust (e.g. being hit with bird droppings); Scare (e.g. a person watching a video, getting a surprise and jumping back); Pain (e.g. a skateboarder falling into a lamppost). Table 39 provides an overview of which stimuli were included within each study, whether the feature of the video was a human or animal, along with a link to the original video which was edited before being shown to participants.

These videos were downloaded and slightly edited, for instance, making them shorter and brighter. They were then uploaded to YouTube where they were privately listed (i.e. would not appear if searched for). The videos were used under fair dealing copyright, specifically for non-commercial research purposes acknowledging the source material (UK Copyright Service, 2009). The videos were subsequently

embedded into a survey hosted via Qualtrics. The questions asked in each study are specified within the method section for each study, along with the scoring method for the stimuli.

The Autism-Spectrum Quotient (AQ)

The AQ was used in many of the pilot studies in 2 formats - the original self-report format (Baron-Cohen et al., 2001) and the adolescent version which is a parental or informant-rater version (Baron-Cohen et al., 2006). These were both marked using the original dichotomous mark scheme. For further, more detailed information on the AQ see Chapter Two.

Table 39: Videos included in their respective studies, alongside links to original videos.

Video Number	Description*	Included in study				Source video
		1	2	3	4	
1	Awkward (H)	X				https://youtu.be/yqqOZPcfAIM
2	Confusion (H)	X				https://youtu.be/yqqOZPcfAIM
3	Awkward (H)	X				https://youtu.be/yqqOZPcfAIM
4	Sour sweet (H)	X				https://youtu.be/ini1EWqTgd4
5	Sour sweet (H)	X				https://youtu.be/ini1EWqTgd4
6	Yawn (H)	X	X	X	X	https://youtu.be/sTvWK7G05PI
7	Scare (H)	X	X			https://youtu.be/_nWMnrn98bc
8	Awkward (H)	X	X			https://youtu.be/aIXh9cXnOYA
9	Cute (H)	X	X	X	X	https://vine.co/v/MJUrli1ie1J
10	Disgust (H)	X				https://youtu.be/suYz5SHFHyA
11	Scare (H)	X	X	X	X	https://youtu.be/eCoyGbOkNQQ
12	Laugh (A)	X	X			https://vine.co/v/enxuWhadEKP
13	Scare (A)	X	X			https://youtu.be/SF3RSxiWd6M
14	Scare (A)	X				https://youtu.be/SF3RSxiWd6M
15	Scare (A)	X	X			https://vine.co/v/enllJ1YizKT
16	Awkward (A)	X				https://youtu.be/T0axoKkiPQQ
17	Cute (A)	X	X	X	X	https://youtu.be/0Bmhjf0rKe8
18	Confusion (A)	X	X			https://tinyurl.com/ya2l2pk3
19	Pain (A)	X				https://youtu.be/rG8v8heFHSY
20	Scare (A)	X				https://youtu.be/SF3RSxiWd6M
21	Scare (A)	X				https://youtu.be/WlRxNSRA7Rg
22	Awkward (A)	X				https://youtu.be/yZ1Vqs_gu1Y
23	Cute (A)	X	X			https://vine.co/v/HJUeli1ij1J
24	Cute (A)	X	X			https://youtu.be/RP4abiHdQpc
25	Pain (H)	X	X			https://youtu.be/zx7CRlYiR2A
26	Shock (H)	X				https://youtu.be/uUwuxxZVANs
27	Cute (H)	X				https://youtu.be/LOJA3pdbFZM

28	Sour (H)	X	X			https://youtu.be/9h5mwoTwDBk
29	Sour (H)	X				https://youtu.be/9h5mwoTwDBk
30	Sour (H)	X				https://youtu.be/9h5mwoTwDBk
31	Sour (H)	X				https://youtu.be/9h5mwoTwDBk
32	Scare (H)	X				https://youtu.be/LOJA3pdbFZM
33	Cute (A)	X				https://vine.co/v/h15HpTJwwJ3
34	Cute (A)	X				https://vine.co/v/hb6MTxrLD6U
35	Scare (H)	X				https://youtu.be/uUwuxxZVANs
36	Laughter (H)	X	X	X	X	https://youtu.be/LOJA3pdbFZM
37	Laughter (A)	X	X	X	X	https://youtu.be/JnemvwqWdhQ
38	Pain (A)	X				https://youtu.be/ltS0G7KA7XQ
39	Scare (A)	X				https://youtu.be/WlRxNSRA7Rg
40	Scare (A)	X	X	X	X	https://youtu.be/EvvqZ01UGSA
41	Pain (H)	X	X	X	X	https://youtu.be/-JdQ00rrdIo
43	Yawn (A)	X				https://youtu.be/B907aaDw7Ec
44	Yawn (H)	X	X			https://youtu.be/sTvWK7G05PI
45	Yawn (A)	X	X	X	X	https://youtu.be/B907aaDw7Ec
46	Yawn (A)	X	X			https://youtu.be/B907aaDw7Ec
47	Pain (A)	X				https://youtu.be/ltS0G7KA7XQ
48	Scare (A)		X			https://youtu.be/_UxoDRHmXfs
49	Disgust (A)		X	X	X	https://youtu.be/MOxNlmNH7l8
50	Scare (A)		X			https://youtu.be/oCzwwKH6Kyk
51	Disgust (H)		X	X	X	https://youtu.be/6gpgZpIp8CA
61	Yawn (H)			X	X	https://youtu.be/OGvFxLREIJE
62	Yawn (H)			X	X	https://youtu.be/1Fs8Tmy2VVo
63	Yawn (H)			X	X	https://youtu.be/0W9XgP3omRA
64	Yawn (H)			X	X	https://youtu.be/LtCX8Rsb38s
65	Yawn (H)			X	X	https://youtu.be/Gk7Pohyqqko

*Note. A human subject in the video is indicated by 'H' and an animal, by 'A.'

5.2.2 Study 1: The initial pilot

5.2.2.1 Brief overview

This pilot study used the stimuli outlined in the general method. The study set out to identify videos that produce consistent reactions and could be used in future work. As this pilot was not about autism screening, no measure of autistic traits was taken. Instead the stimuli that were identified as producing consistent reactions, were retained for the future studies in which participants provided measures of autistic traits.

5.2.2.2 Aims and hypothesis

The aim of the study was to identify which videos produce consistent self-reported reactions. It was hypothesised that a number of the videos would produce these, but no hypothesis about which videos may or may not produce these was made.

5.2.2.3 Participants

Participants were an opportunity sample recruited via word of mouth and social media. The final sample comprised 10 participants (*M* age = 29.90, *SD* = 15.24; male = 3). The education levels of the participants were as follows: 1 person had no formal qualifications, 1 had a vocational qualification, 6 had higher leavers certificates (e.g. A-level), 1 had an undergraduate degree, and 1 had a postgraduate degree. No one in the sample reported any diagnoses (ergo, a non-autistic sample). Two of the 10 participants only partially completed the study, while the remaining 8 completed the study in its entirety. Exclusion criteria for the study were: anyone under 18 years old; a severe, non-corrected visual impairment.

5.2.2.4 Stimuli

A total of 46 videos were used (see Table 39 for a summary of their content, and links to the original videos). The videos were showed to participants through Qualtrics using a YouTube plugin. For each video that the participant saw, they were asked ‘What was your reaction to this video?’ and were given a response box to write their response in.

5.2.2.5 **Procedure**

The study was approved by Northumbria University's Health and Life Sciences ethics committee. Participants were directed to the online questionnaire where they were presented with detailed information about the study, and an email contact address to direct any questions to. Once they consented to take part in the study, they were asked to provide demographic information (age, gender, education, any relevant diagnoses), and were presented with the videos and asked the questions described in the 'stimuli' section.

5.2.2.6 **Analysis strategy**

Written responses were read by the researcher. Longer responses were summarised while retaining the core message behind the person's response. The responses and/or summaries were read together, and a brief summary of the responses was written. A decision was then made by the main researcher and their supervisor about whether these responses were consistent across the group and what the video was expected to elicit. The original responses were also referred to at this point, if necessary. Videos which indicated inconsistent responses were highlighted for exclusion from further studies, whereas those which elicited consistent responses were identified for use in future studies.

5.2.2.7 **Results**

The summaries and decisions can be seen in Table 40. Of the 47 videos shown to participants, 21 were highlighted as eliciting consistent responses and were retained.

Table 40: Summaries of text responses in Study 1.

Video Number	Description*	Summary of responses**	Consistent and retained
1	Awkward (H)	Some awkward, some laughter	
2	Confusion (H)	Laughter, lots of response of 'nothing'	
3	Awkward (H)	Laughter, lots of response of 'nothing'	
4	Sour sweet (H)	Some cute, some smile, some laughter, some 'screwing up face'	
5	Sour sweet (H)	Some funny, some smiles, some 'nothing'	
6	Yawn (H)	Yawns and not funny	Yes
7	Scare (H)	Shock or the expectation for something to jump out, some laughter	Yes
8	Awkward (H)	Awkward, embarrassment, 'nothing'	Yes
9	Cute (H)	Cute, funny, laughter	Yes
10	Disgust (H)	Empathy, dislike of birds, amusement	
11	Scare (H)	Shock, jumping, amusement, empathy	Yes
12	Laugh (A)	Amusement	Yes
13	Scare (A)	Surprise, amusement	Yes
14	Scare (A)	Amusement, feeling sorry for the animal, one jump	
15	Scare (A)	Mostly amused	Yes
16	Awkward (A)	Mostly no real reaction	
17	Cute (A)	Cute, funny	Yes
18	Confusion (A)	Amusement	Yes
19	Pain (A)	Feeling sorry for the animal, finding it cruel	
20	Scare (A)	Shock, amusement, sadness, concern	
21	Scare (A)	Feeling sorry for the animal, amusement	
22	Awkward (A)	Mostly no real reaction	
23	Cute (A)	Cute, amusement	Yes
24	Cute (A)	Amusement, cute, smile	Yes
25	Pain (H)	'Ow', noting that the subject is in pain	Yes
26	Shock (H)	Jump, annoyed at the person doing the scaring, 'nothing'	
27	Cute (H)	Laughter, some cute	
28	Sour (H)	Cruel, screwing up face, laughter	Yes
29	Sour (H)	Amusement, cringing, screwing up face, nothing	
30	Sour (H)	Amusement, screwing up face, 'nothing'	
31	Sour (H)	Amusement and 'nothing'	
32	Scare (H)	Anger at person doing scare, amusement	
33	Cute (A)	Cute, amusement, 'nothing'	
34	Cute (A)	Cute, 'nothing'	
35	Scare (H)	Amusement, jumping, anger toward person doing the scare	
36	Laughter (H)	Laughter, cute	Yes

37	Laughter (A)	Laughter	Yes
38	Pain (A)	Sympathy for the animal	
39	Scare (A)	Cruel, annoyed at the person doing the scare	
40	Scare (A)	Sympathy, tension, jumped	Yes
41	Pain (H)	Pain, sympathy	Yes
43	Yawn (A)	Cute, 'nothing', yawn	
44	Yawn (H)	Yawns, 'nothing'	Yes
45	Yawn (A)	Cute, yawning, 'nothing'	Yes
46	Yawn (A)	Yawning, 'nothing'	Yes
47	Pain (A)	'Nothing', amusement	

*Note. A human subject in the video is indicated by 'H' and an animal, by 'A.'

**Note. Where 'not funny' was seen in the original data, this is understood as the person was not feeling/doing anything in response to this video

5.2.2.8 **Discussion**

This study showed that videos, which feature both animals and humans that produce relatively consistent reactions, can be identified. The study identified 21 videos for use in more robust, future studies. The videos with more consistent responses are important for the subsequent studies discussed in this chapter. These will facilitate the investigation of mimicry and emotional contagion, and whether they relate to autism in such a way that videos, such as those identified in the pilot study, can be used in lieu of a more traditional screening tool. Consistent responses to the retained videos by typically developing participants were important, as this indicated that the stimuli may be facilitating mimicry (if differences in reactions exist, then the viewers were not mimicking what they saw). In addition, having consistent responses allows for a scoring system to be developed in future studies. This was important, as having a clear, simple scoring procedure can reduce errors (Glascoe, 2005).

There were drawbacks to this pilot study which should be acknowledged. First, the sample was small, with only 10 participants, 2 of whom did not complete the questionnaire fully. This leads to the second issue, study length. Feedback from participants indicated that the 2 who did not finish found the task to be too long and unengaging; the final task should reflect this by being both shorter and more engaging. Finally, no formal mark scheme was employed here, instead participants' responses were simply summarised by the researcher. While this allowed for a good overview and an initial insight into the

types of responses, this is not a robust enough method to be used with a final screening tool and accordingly, a different approach will be taken with future work.

Overall, while there were issues with the methodology employed in this pilot study, it was a good first step into investigating the possibility of stimuli that prompt responses based on mimicry and emotional contagion, being included in behavioural screening tools in the future. It also provided an initial set of stimuli to work with in future studies.

5.2.3 **Study 2: Identifying videos that yield strong reactions**

5.2.3.1 **Brief overview**

This study used the 21 videos identified in the initial pilot study, as well as 4 additional videos; 2 to prompt scare reactions and 2 to prompt disgust reactions. These additional videos were included as although scare videos in Study 1 were identified as producing relatively consistent responses, these were often of amusement rather than fear. In addition, the disgust videos used in the pilot did not produce consistent reactions. The 4 new videos were included as they were considered to be better stimuli than the originals.

This study set out to reduce the number of stimuli further, in order to ensure a more refined set of stimuli and balance the set between animals and humans. As outlined previously, literature indicates that people generally empathise more with humans (Franklin et al., 2013), or more human-like animal stimuli (Westbury & Neumann, 2008), and that autistic people benefit from the presence of animals (O'Haire, 2017). Due to this, both animal and human stimuli were included in Study 2, as it may have allowed for a greater range of reactions from participants. As with the pilot study, participants self-reported their reactions by typing them, though in this study a mark scheme was also used to keep analysis and scoring as consistent as possible. This mark scheme was based on whether or not the person's response indicated that mimicry or contagion had taken place. A later study (see 5.2.7) evaluated the interrater reliability of this marking scheme.

During this study, participants also completed the AQ. In a later analysis (see 5.3.5), results from this study will be combined with those from Study 3 to investigate whether responses to videos relate to AQ scores.

5.2.3.2 **Aim and hypothesis**

The aim of this study was to identify a smaller set of stimuli that include both animal and human videos which are capable of producing strong, consistent reactions, and that can be used in a later study linking the reactions to autistic traits. It was hypothesised that for each type of stimuli (e.g. Laughter) 1 video of a human and 1 video of an animal would show a stronger reaction than the alternatives and would, therefore, be identified as stimuli to use in future research.

5.2.3.3 **Participants**

Participants were an opportunity sample recruited by word of mouth, through social media, and the university participant pool. The latter rewards student participants with credits that, at a later time they are able to pass onto other people as a reward for taking part in their own research. In total, 45 participants were recruited (*M* age = 21.56, *SD* = 5.68; male = 11). In terms of education, the sample had the following qualifications: 1 person with no formal qualifications, 1 school leaver (e.g. GCSE), 1 with vocational qualifications, 40 higher school leavers (e.g. A-Level), and 2 with undergraduate degrees. One person in the sample reported a learning difficulty (i.e. dyslexia), 2 reported other conditions (cognitive impairment and prosopagnosia), and 42 reported no diagnosed conditions. Exclusion criteria for the study were: anyone under 18 years old; a severe, non-corrected visual impairment.

5.2.3.4 **Stimuli**

As outlined in the brief overview, this study employed the 21 items retained in Study 1 as well as 4 new videos (2 x Disgust; 2 x Fear). The new stimuli were felt to be of better quality, in terms of picture and audio, and therefore more likely to elicit mimicry reactions from participants.

As in Study 1, the videos were displayed to participants through Qualtrics using a YouTube plugin. Participants were asked ‘What was your reaction to this video?’ and were given a response box to write their response in. Additionally, participants were asked to rate the strength of their reaction on a sliding

scale from 0 to 10, with 0 being no reaction and 10 being a strong reaction. Participants also completed the AQ as part of the online task.

5.2.3.5 **Procedure**

The study was approved by Northumbria University's Health and Life Sciences ethics committee. Participants were directed to the online questionnaire where they were provided with detailed information about the study, and a contact email address to ask any questions that they may have. Once consent to take part had been given, they were asked to provide relevant demographic information (age, gender, education, occupation, diagnoses). Then, participants were presented with the materials outlined in the stimuli section.

5.2.3.6 **Analysis strategy**

Responses to video were coded as either the 'expected' response (1), 'unexpected' response (2), or 'missing' (3). For example, if the video was a person laughing, the 'expected' response could include the participant laughing, smiling, feeling happy, or similar. Other responses such as 'nothing' would be classed as 'unexpected', no response left was considered 'missing' (for scoring criteria see Table 41). The interrater reliability of this mark scheme is explored in section 5.2.7 (Studies 2 – 5 Interrater reliability).

Descriptive statistics were calculated for participant responses and the reported strength of these reactions. For each type of video (e.g. Yawn), the video with the strongest reaction featuring a human subject, and the video with the strongest reaction featuring an animal subject were retained for use in the next study. This would help ensure that the videos being used as stimuli were sufficient to produce a reaction. While the responses were scored using the aforementioned mark scheme, this was for use in a later analysis and only the descriptives will be provided here. This analysis focussed on the reported strength of reaction.

Table 41: Marking criteria for Studies 2, 3, and 5.

Expected reaction	Examples of accepted responses (i.e. scored '1')
Awkward	Feelings of awkwardness, discomfort, and intrigue or interest
Confusion	Feelings of confusion, not understanding, and finding it strange
Cute	Responding that the video is cute, saying 'aww,' smiling, and feelings of happiness
Disgust	Feelings of disgust, sick, turning up of nose, and feeling ill
Laughter	Laughing, feeling happy, and smiling
Pain	Wincing, saying 'ouch'/'oww' jolting, and feelings of discomfort
Scare	Feeling scared, jumping, and getting a fright
Sour	Tightening of the mouth, 'screwing up' of face, wincing, and feeling sorry for the person in the video
Yawn	Yawning, feelings of tiredness, and trying to not yawn

5.2.3.7 **Results**

Descriptive statistics show the percentage of people reacting in a manner that demonstrates mimicry or contagion (as scored by the main researcher) and self-reported strength reactions (see Table 42). The videos with the highest reported contagion responses were the Human Disgust (video 51; 93.3%) and Animal Disgust (video 49; 91.1%). Conversely, the videos with the lowest contagion responses were Animal Scare (video 15; 4.5%) and Human Sour (6.7%). The videos for humans and animals of each type (e.g. Scare, Disgust) with the highest strength reactions were identified for use in future studies. For instance, for Animal Cute, video 17 was chosen rather than videos 23 or 24 on account of having a higher rated score.

Due to inconsistent responses in Study 1, only 1 Confusion video was included in Study 2. This had a low rate of 'expected' reaction (9.1%) and was not included in subsequent studies. Similarly, no videos of Animal Pain were included in Study 2, due to prior low rates of reaction, so only video 41 was retained which showed a Human Pain response. The retained videos were as follows: for Yawning 45 and 6, for Scare 40 and 11, for Cute 17 and 9, for Pain 41, for Disgust 49 and 51, and for Laughter 36 and 37. The identified videos are marked with an asterisk in Table 42.

Table 42: Descriptive statistics for Studies 2 and 3, broken down by video.

Video Number	Description*	Study 2		Study 3	
		'Expected' %	Mean *(SD)* strength	'Expected' %	Mean *(SD)* strength
6	Yawn (H)	52.30	4.00 (3.20)*	25.30	3.36 (2.92)
7	Scare (H)	34.10	3.22 (1.97)		
8	Awkward (H)	25.00	3.56 (2.28)		
9	Cute (H)	93.20	4.50 (2.26)*	87.50	5.26 (2.57)
11	Scare (H)	76.70	4.86 (2.59)*	56.80	5.19 (2.91)
12	Laugh (A)	81.80	4.86 (2.18)		
13	Scare (A)	15.90	3.44 (2.06)		
15	Scare (A)	4.50	4.12 (2.43)		
17	Cute (A)	90.90	5.40 (2.23)*	92.60	5.79 (2.38)
18	Confusion (A)	9.10	5.40 (2.58)		
23	Cute (A)	71.10	3.59 (1.83)		
24	Cute (A)	88.90	5.33 (2.29)		
25	Pain (H)	43.20	5.02 (1.97)		
28	Sour (H)	6.70	5.02 (2.07)		
36	Laughter (H)	82.20	5.55 (2.55)*	86.30	6.14 (2.56)
37	Laughter (A)	73.30	4.53 (2.60)*	74.70	5.48 (2.48)
40	Scare (A)	45.50	4.41 (2.18)*	49.50	4.70 (2.65)
41	Pain (H)	82.20	5.02 (2.05)*	75.30	5.47 (2.37)
44	Yawn (H)	48.90	3.41 (2.75)		
45	Yawn (A)	9.10	3.63 (2.58)*	7.40	4.18 (2.73)
46	Yawn (A)	18.60	2.89 (2.27)		
48	Scare (A)	61.40	3.95 (2.28)		
49	Disgust (A)	91.10	6.36 (2.39)	83.50	6.53 (2.87)
50	Scare (A)	45.50	3.17 (1.66)		
51	Disgust (H)	93.30	6.39 (2.39)	83.30	6.03 (2.64)
61	Yawn (H)			12.00	2.13 (2.61)
62	Yawn (H)			21.30	2.50 (2.79)
63	Yawn (H)			26.40	2.91 (2.92)
64	Yawn (H)			13.00	1.97 (2.42)
65	Yawn (H)			26.70	2.25 (2.52)

*Note. A human subject in the video is indicated by 'H' and an animal, by 'A.'

5.2.3.8 **Discussion**

The results here further refined the potential stimuli set for use as a behavioural screen, by ensuring that only stimuli which produced relatively consistent reactions and provoked the strongest reactions will be included in future studies. It is hoped that this will make such reactions easier to observe and/or the participant to self-report, which will be important when creating a tool for use with people with learning disability. A challenge that many people with learning disability face is the interpretation and expression of emotion, then subsequently relaying those feelings to others (Mcclure, Halpern, Wolper, & Donahue, 2009; Sovner & Hurley, 1986). Hopefully, by using stimuli that elicit strong reactions, the task of reporting emotions and feelings will be made easier. This study also identified videos of both animals and humans that elicited the strongest reactions. This is because people seemingly empathise differently with animals compared to humans (Westbury & Neumann, 2008).

This study had limitations which must be addressed in subsequent studies, if a robust screening tool is to be created. First, the analysis was principally driven by the strength of participants' reactions rather than the ability of the stimuli to induce mimicry. The videos with the highest strength reactions which were retained also largely produced consistent reactions, although the 2 retained Yawn stimuli have quite low expected responses (Human = 52.30%; Animal = 9.10%). The reason that strength was chosen in this study was that the reactions needed to be strong enough to be recognised and self-reported by participants. Secondarily, if only the stimuli with the most consistent reactions were retained, this would leave very little variation in the responses and would be unlikely to result in an effective screening tool.

A further, important limitation of this study was that it used self-report rather than observational data, whereas the ultimate aim of the research is to develop an observation based behavioural screening tool.

In conclusion, this study refined the stimuli set further to ensure it was both shorter and able to evoke strong reactions in people. The data collected here, when combined with further use of this stimuli set and the mark scheme used, will provide an insight into whether mimicry can provide an indication about a person's AQ score. In turn, this would mean reactions to mimicry could be utilised as an alternative to more conventional screening tools.

5.2.4 **Study 3: Using a refined stimuli set**

5.2.4.1 **Brief overview**

This study used a more refined and smaller stimuli set than in previous studies, using the stimuli identified as evoking strong participant reactions in Study 2, with an additional 5 Yawn videos. The study allowed further exploration of self-reported reactions, and when combined with the data from Study 2, gave an indication of whether these relate to a person's self-reported AQ score. This study also expanded recruitment beyond non-autistic samples and includes participants who report a diagnosis of autism. The addition of 5 videos of yawning was to investigate whether a 'yawn threshold' was in some way related to autism. Previous literature indicates limited contagious yawning is associated with autism (e.g. Senju et al., 2009). The idea here was that if a participant does not yawn when viewing the first yawn, it may not be indicative of autism, but not contagiously yawning after a number of videos of yawning may indicate increased autistic traits. This study aimed to investigate whether yawning after being exposed to a different number of videos showing yawning was meaningful in the detection of autism or autistic traits.

5.2.4.2 **Aim and hypothesis**

The first aim of this study was to collect more data which could be combined with data from Study 2, for analysis in Study 4. The second aim of this study was to investigate a 'yawn threshold' which is whether the number of yawns viewed before the person yawns themselves is related to their AQ score. For this second aim, it was hypothesised that a potential indicator for higher AQ scores may be the need to observe multiple videos of yawning before a contagious yawn response was evoked.

5.2.4.3 **Participants**

Participants were an opportunity sample recruited by word of mouth, via social media, the university participant pool (see 5.2.3.3), and a charity that supports autistic people. Ninety-one participants were recruited (*M* age = 34.78, *SD* = 13.84; male = 21). The following qualifications were reported: 2 held no formal qualifications, 4 were school leavers (e.g. GCSE), 1 held vocational qualifications, 18 were higher school leavers (e.g. A-Level), 46 had undergraduate degrees, 23 had postgraduate degrees and 3

held other qualifications. Eighteen participants reported a diagnosis of autism, 7 a learning difficulty (e.g. dyslexia), 3 other (e.g. anxiety), and 74 reported no diagnosis. Exclusion criteria for the study were: anyone under 18 years old; a severe, non-corrected visual impairment.

5.2.4.4 **Stimuli**

The highest strength items in terms of elicited response which were identified in Study 2 were used in this study. Five additional videos of different 'actors' yawning (volunteer academic staff and postgraduate students) were also included. These were filmed against a plain background, and mainly focussed on the person's face and chest. The 'actors' in the videos were asked to yawn. If they were unable to do so, they were asked to think about yawning and to watch the researcher yawning. The camera was left recording for the whole session, but only the part where the person yawned was used in the study. All videos lasted less than ten seconds. Links to Yawn videos are available in Table 39.

As in Studies 1 and 2, Qualtrics was used to provide participant information, record consent, and display the videos to participants via a YouTube plugin. Participants were asked 'What was your reaction to this video?' and to rate the strength of their reaction on a scale of 0 to 10, with 0 being no reaction and 10 being a strong reaction. Participants also completed the AQ (Baron-Cohen et al., 2001).

5.2.4.5 **Procedure**

The study was approved by Northumbria University's Health and Life Sciences ethics committee. The procedure was identical to Study 2.

5.2.4.6 **Analysis strategy**

As in Study 2, responses to videos were coded as 'expected' (1), 'unexpected' (2), or 'missing' (3). Descriptive statistics were calculated for responses and self-reported strength scores. As with Study 2, a second rater coded a subset of the sample responses, the results of which will be explored in section 5.2.7. It was intended that the responses to the videos of yawning would be viewed separately, to allow the point at which the person first yawned to be identified (e.g. after seeing the second yawn, the fourth yawn, etc.) Unfortunately, unknown to the main researcher, the randomise function in Qualtrics did not

collect information about the order that the stimuli were viewed in. As a result, this analysis could not be conducted.

5.2.4.7 **Results**

Descriptive statistics were calculated for both the percentage of expected responses and the self-reported strength of responses to the videos, see

Table 42 (displayed in Study 2) for descriptive statistics. The descriptives show that videos 49 and 51, the Animal and Human Disgust stimuli, have a lower expected response rate in this study as compared to Study 2. The highest rate of expected responses was found to be for video 17, the Animal Cute stimuli.

It should be noted here that the Yawn video which previously evoked the highest percentage of responses in Study 2 (video 6), saw a considerably lower percentage of participants showing expected reactions in Study 3. Additionally, the reported strength of reaction to the video dropped from 4.00 to 3.36. The additional five Yawn videos (videos 61-65) all showed expected reactions of under 30% (see Table 42 for details). Overall, the mean expected Yawn response was higher in Study 2 (M = 32.23%; SD = 21.61) than in Study 3 (M = 18.67%; SD = 7.94). No test of difference was calculated here due to the differences between samples (Study 3 included autistic adults).

5.2.4.8 **Discussion**

This study gathered additional data about the stimuli and included autistic participants. More examples of yawn stimuli were included in this study. Perhaps due to this, the congruency of the reactions to the yawn stimuli appeared lower as compared with previous studies. This difference was potentially due to the repeated exposure, leading to a fatigue effect; much like Provine's (1992) finding that people are less likely to respond positively to laugh stimuli after repeated exposure due to irritation. The impact of this on the final behavioural screening tool will be discussed further in the next section and the general discussion.

This study had a limitation common to the other studies already discussed in this chapter: the reliance on self-reported reactions rather than observing whether mimicry occurred. This means that the participant may not be reporting their reaction accurately.

The data from this study will be combined with the data from Study 2 in the next section, which will provide more meaningful analysis on the phenomenon of mimicry and emotional contagion.

5.2.5 Study 4: Combining data for machine learning

5.2.5.1 Brief overview

Machine learning can provide insights about data that more frequently used analyses cannot. The approach estimates a person's score on one variable from their responses on a set of others - in this case, to predict a person's AQ score from whether they responded to a video in a manner that indicates mimicry. In order to create a larger, more robust dataset, data common to both Studies 2 and 3 were combined and used in this study. This included: gender, age, education, video 6 (Yawn), video 9 (Cute), video 11 (Scare), video 17 (Cute), video 36 (Laughter), video 37 (Laughter), video 40 (Scare), video 41 (Pain), video 45 (Yawn), video 49 (Disgust), and video 51 (Disgust). The source of the data (i.e. Study 2 or Study 3) was also used.

5.2.5.2 Analysis Strategy

As noted above, only data from videos present in both Studies 2 and 3 were used in the analysis. The scoring used in Studies 2 and 3 (see Table 41) was used here, and the AQ scores from participants were calculated using the original dichotomous scoring format.

A bottom-up approach using machine learning implemented in cforest (Hothorn, Hornik, Strobl, & Zeileis, 2010) in R (R Core Team, 2013) was employed. This method used an extension of Random Forests, whereby 10,000 regression/classification trees were created to discover patterns within the data (Hastie, Tibshirani, & Friedman, 2009; Hothorn et al., 2010; Strobl, Malley, & Tutz, 2009). This algorithmic approach can handle correlated data, interactions between variables, and non-linear patterns within the data, and implements multiple splits along the same variable. Further, this specific approach also corrects for multiple testing, and prevents variables with greater numbers of outcomes arbitrarily being highlighted as highly predictive. As this study was exploratory and no specific hypotheses were outlined, this approach was extremely valuable.

Subsequent to the generation of the 10,000 trees, variable importance can be calculated (Strobl et al., 2009). Variable importance is based upon permuting a predictor variable, therefore when genuinely predictive variables are shuffled out this will lead to a notably worse prediction (Janitza, Strobl, & Boulesteix, 2013; Strobl et al., 2009). This yields variable importance, which identifies variables that have and do not have predictive ability.

Analyses were run in R 3.4.1 (R Core Team, 2013). Additional information can be found in ESM (Chapter 5 / S4).

5.2.5.3 **Participants**

Participants were sourced from Study 2 ($N = 45$) and Study 3 ($N = 97$; note that there are 6 additional participants here, this analysis is able to process cases with missing data provided that it is not the dependent variable). When combined, this totalled 142 participants of whom 32 were male and 110 were female, with a mean age of 30.56 ($SD = 13.26$).

5.2.5.4 **Results**

Descriptive statistics were calculated for all items, these can be found in Table 43. AQ scores were calculated, finding a mean score of 19.95 ($SD = 10.31$), which is notably lower than Baron-Cohen et al.'s (2001) cut-point of 32; where people scoring 32 or higher should be considered for further assessment.

Table 43: Descriptive statistics from Studies 2 and 3, combined data.

Video Number	Description*	'Expected' %
6	Yawn (H)	33.10
9	Cute (H)	88.00
11	Scare (H)	61.30
17	Cute (A)	90.10
36	Laughter (H)	83.80
37	Laughter (A)	73.20
40	Scare (A)	47.20
41	Pain (H)	22.50
45	Yawn (A)	7.70
49	Disgust (A)	14.10
51	Disgust (H)	85.90

*Note. A human subject in the video is indicated by 'H' and an animal, by 'A.'

The random forest plots were calculated using 10,000 trees. Each tree could individually consider a (random) combination of 5 variables at any split point ('mtry'). At the splits, the strongest association was selected and used in that tree. The number of factors was derived from the number of predictor variables (15) divided by 3, as this can make accurate predictions about the data. Unlike computing all possible combinations, this approach is not prohibited by computational power (Cutler et al., 2007; Liaw & Wiener, 2002).

When combined, the random forests analysis revealed that it is possible to use participants' responses to these videos to predict their AQ score. The variables can be plotted in order of importance to predict said AQ score. Variables of increasing importance are displayed from left to right while the dashed line indicates significance (see Figure 4). The variables with the highest importance were 41 (Human Pain) and 37 (Animal Laughter). Notably both Scare videos were of least importance (videos 40 and 11), both Disgust videos were of little importance (videos 49 and 51), and the Yawn video (17) was below the significance line. Implications of this latter point will be discussed. There was a general trend across the videos for the human videos to be deemed of higher significance than the animal videos.

As random forests are truly random, the analysis was re-run, with a different random starting seed. The ranking of variable importance between these analyses was perfect ($r_s = 1.00$, $p < .001$). The second analysis can be found in the ESM (Chapter 5 / S4 / ML & dotplot2).

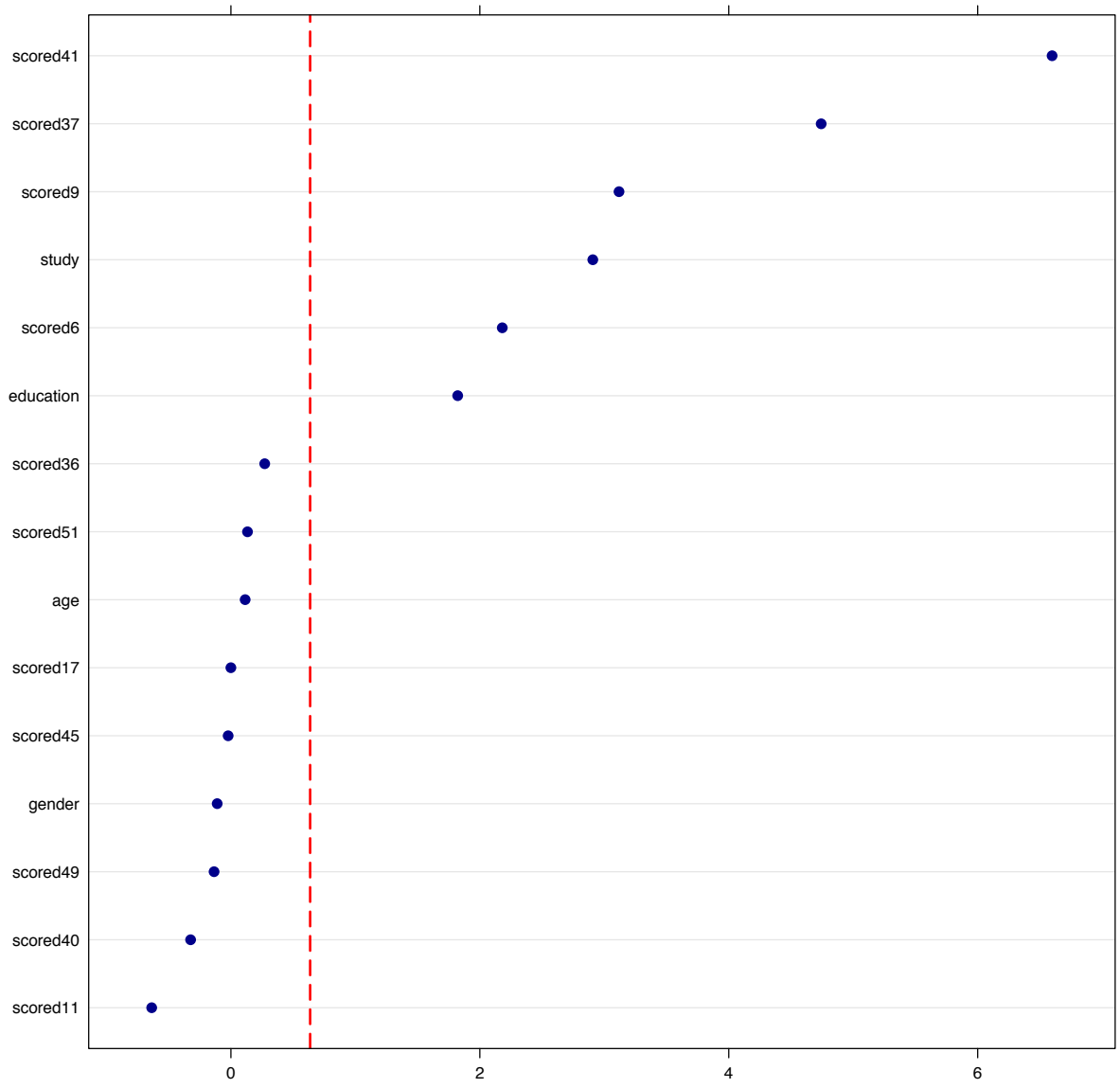


Figure 4: Dot plot showing variable importance. Dashed vertical line indicates significance, predictors to the right are significant (the benchmark for significance is the (absolute) value importance of the worst predictor in the set).

The Bland Altman plots indicate that the model largely accurately predicted AQ scores (i.e. most points between the control lines (1.96*SD)), however, for the lower range it tended to underestimate and for the upper range it tended to overestimate. There were around 5 cases for which the model severely overestimates, these were extreme AQ scores (> 30+), see Figure 5. A correlation which was run between actual and predicted values indicated a significant relationship, $r_s = .49$, $p < .001$.

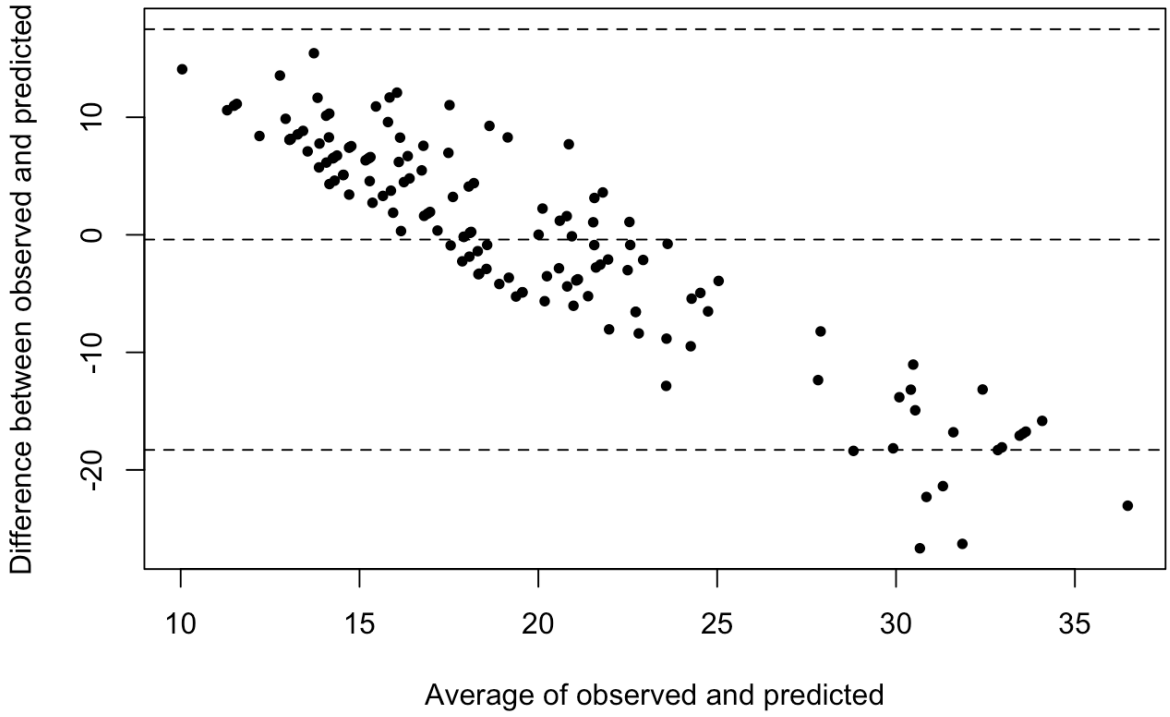


Figure 5 Bland Altman plot: the relationship between the average of observed and predicted values, and the difference between observed and predicted scores.

5.2.5.5 **Discussion**

Using the combined data from Studies 2 and 3, machine learning illustrated that it is possible to make predictions about an individual's AQ score from their self-reported mimicry and emotional contagion reactions. Previous research shows that when simply viewing stimuli, without any prompts to react, a diagnosis of autism is related to whether someone is likely to mimic the emotional state that they are viewing. Specifically, that being autistic reduces the chance of mimicking a viewed stimulus (e.g.

McIntosh et al., 2006; Yoshimura et al., 2015). Other research has investigated responses to yawning using a similar underlying premise, for example, studies find that autistic people are less likely to contagiously yawn compared to non-autistic control participants (Helt et al., 2010; Senju et al., 2007). Study 5 did not use one particular stimuli type, instead it combined many types of stimuli that can be considered contagious (e.g. something someone may mimic), or empathy-based stimuli. Rather than comparing people with and without autism, it used participants' AQ scores as a self-reported measure of autistic traits. This allowed the sample to be tested on a continuum rather than discreet groups, which as there were quite low numbers of autistic people in the sample, allowed for a stronger analysis. In past research (e.g. Chan & Tseng, 2017), using AQ score rather than diagnosis has led to null results. The possible difference here was that because multiple types of stimuli were utilised, there was a greater heterogeneity of responses. Additionally, this sample included people who reported being autistic, which again creates a greater degree of heterogeneity compared to past research (e.g. Chan & Tseng, 2017). The finding that it is possible to estimate AQ score from self-reported reactions to video stimuli is important, as it raises the possibility that reactions to video stimuli can act as an indicator of whether or not the person is likely to be autistic.

It is important to express caution about these findings. The data and analysis were based on AQ scores, rather than on a person's diagnosis. While the two are related, with higher scores indicating a higher chance of the person being autistic, it is not necessarily always the case, as even high scorers may not actually be autistic (Baron-Cohen et al., 2001). In addition, previous research on mimicry and contagious yawning has found diagnosis of autism to be relevant (Helt et al., 2010; Senju et al., 2007), rather than a person's AQ score (see: Chan & Tseng, 2017). This may still be the case in the instance of yawning, as the machine learning used in the present study indicated that the reactions to yawns were some of the least significant in the sample.

There are two further reasons for caution: firstly, video quality and secondly, differences between the stimuli that the person has been exposed to. On the former matter, the videos used were pre-existing for the most part, having been sourced from internet videos and edited to be relevant to the study. On the one hand, this is advantageous as they are naturalistic but on the other, they are not well controlled, and

aspects including the background, length of the video, and quality of the video fluctuate between stimuli. This may mean, simply put, that better quality videos could have led to stronger reactions, and thus may count as more influential on the machine learning outcome. On the latter issue, the data here was compiled from 2 studies because there was a large amount of overlapping data and the methodologies were very similar. While similar, the methodologies were not identical, and the overall nature and number of videos that people viewed in each study were different. For instance, while Study 2 included 4 Yawn videos and Study 3 included 7, only 2 of these featured in both studies (videos 6 and 45). The overall context in which each video was viewed may, therefore, have influenced the responses. For example, as discussed in section 5.2.4, there was a lower rate of contagious yawning in Study 2 than in Study 3, which may have happened because people were exposed to a greater number of yawning videos in the latter study, and inhibition may have occurred. These limitations mean that specific findings, such as that the reactions to the videos of pain are the most significant in estimating AQ scores, should be viewed with caution because the results may differ according to the context within which a specific video was viewed.

As can be seen from the Bland Altman plot, the forest plots are only able to estimate a person's AQ score within about 10 points and it becomes notably more difficult for those with either extremely high or low scores. These more difficult estimates could be overcome with a larger sample as it would likely produce a more accurate model, with more robust results. For this reason, the current screening tool is not a suitable alternative to existing screening tools. Despite this, the study shows that with refinement, tools based upon reactions to dynamic stimuli may be able to estimate AQ scores and act as a reliable alternative when using the AQ is not possible. Accordingly, these findings will be built upon in subsequent studies.

5.2.6 **Study 5: Adolescents and observer reporting**

5.2.6.1 **Brief overview**

All of the previous studies in this chapter were conducted with adults. This study extends the research to a group of adolescents, some of whom have diagnoses of autism and learning disability. The study was a small pilot designed to explore whether the procedure would be appropriate for younger people in

addition to adults. This study included observer ratings as well as self-report, enabling comparison between the two methods. In order to explore whether the method can be used by a naïve rater, the observations were carried out by an undergraduate student who had no experience of working with autistic people or screening for autism, rather than by the main researcher. For a screening tool to be effective, it must be able to be used by a wide range of people, with varying levels of knowledge and experience. As was seen in Chapter 4: many people working with autistic people and people with learning disability, have no experience of using screening tools. This study also looked at the more practical aspects of implementing a tool when delivered by a person with no prior knowledge of the area.

5.2.6.2 **Participants**

Participants were recruited via a high school/sixth form in the North East of England. A total of 14 children took part in the study (*M* age = 14.57, *SD* = 2.03; male = 7). Nine children were reported to have no conditions, the remaining 5 were reported to have one or more of the following: 4 were diagnosed with autism, 4 with learning disability, 2 with learning difficulty, 1 with cerebral palsy, 1 with a speech condition, and 1 with an unknown additional need. Thirteen participants were White British and 1 was reported as 'other'. Exclusion criteria for the study were: anyone under 5 years old; a severe, non-corrected visual impairment. Written informed consent was provided by a parent or carer, if under 18 years old.

5.2.6.3 **Stimuli**

This study used the same video stimuli as in Study 3, and the adolescent version of the AQ (Baron-Cohen et al., 2006) was used. This is a parental report rather than a self-report version.

5.2.6.4 **Procedure**

The study was approved by Northumbria University's Health and Life Sciences ethics committee. Participants were recruited through the school. Packs containing participant information, consent forms, and questionnaires were sent home with the children. The pack also included envelopes to return the information to the researchers via school staff. The questionnaires were demographic information sheets (age, gender, relevant diagnoses), and a parental report version of the AQ. The participating children

viewed the stimuli on a computer at school, using the same online survey and questions as described in Study 3, in the presence of the undergraduate student (henceforth referred to as 'the observer'). They were given the option of either typing the responses themselves, as in the previous studies, or verbally stating how they felt and having the observer type the response. The observer also noted on paper, any visible reactions to the stimuli.

5.2.6.5 **Analysis strategy**

This was principally an interrater reliability exercise. Both the participants' self-report reactions and notes taken by the observer were scored using the same method as in Studies 2 and 3 (i.e. indicating whether a person reacted in a manner indicating mimicry or not). However, the observer simply stated whether they felt the person mimicked what they had viewed and indicated either yes or no. The observer's notes were coded by the main researcher, followed by the participant's self-reported reaction, using the same dichotomous scoring method. Kappa values were calculated between the different scores and interpreted according to McHugh (2012).

5.2.6.6 **Results**

The mean AQ score for all participants was 17.00 (*SD* = 6.08). For those who reported a diagnosis of autism, the mean AQ score was 20.75 (*SD* = 3.04) and for those who reported no diagnosis of autism, it was 15.50 (*SD* = 6.40). Descriptive statistics in Table 44 show the expected reactions according to the observer, the observer's notes coded by the researcher, and participants' self-reported reactions also scored by the researcher.

Table 44: Responses of participants by the observer, the observer's notes coded by the researcher or the participant's self-reported reactions coded by the observer.

Video Number	Description*	Expected %		
		Observer	Notes	Self-report
6	Yawn (H)	14.30	14.30	42.90
9	Cute (H)	78.60	85.70	92.90
11	Scare (H)	85.70	21.40	14.30
17	Cute (A)	85.70	85.70	92.90
36	Laughter (H)	92.90	92.90	92.90
37	Laughter (A)	92.90	92.90	100.00
40	Scare (A)	71.40	42.90	71.40
41	Pain (H)	57.10	50.00	57.10
45	Yawn (A)	14.30	14.30	21.40
49	Disgust (A)	85.70	78.60	71.4
51	Disgust (H)	28.60	28.60	42.90
61	Yawn (H)	7.10	7.10	28.60
62	Yawn (H)	21.40	14.30	50.00
63	Yawn (H)	14.30	14.30	50.00
64	Yawn (H)	100.00	100.00	35.70
65	Yawn (H)	14.30	14.30	35.70

*Note. A human subject in the video is indicated by 'H' and an animal, by 'A.'

Observer vs. self-report

Agreement was calculated between the observer and the person's self-reported reaction to the video, as scored by the main researcher. Cohen's Kappa indicated a significant, but weak agreement, $K = .484$, $p < .001$.

Observer's coded notes

Cohen's Kappa showed a significant, moderate agreement, $K = .766$, $p < .001$, between the observer's first-hand rating and scoring of the same observer's descriptive notes by the main researcher. Cohen's Kappa revealed a significant but weak agreement, $K = .054$, $p < .001$ between those notes scored by the main researcher and the person's self-reported reactions.

5.2.6.7 **Discussion**

The results showed that there was substantial agreement (Landis & Koch, 1977) between the observer's ratings of participant responses and ratings of the observer's notes by the researcher, who did not see the participants. A lower, moderate agreement was found between the observer's rating and participants' self-reported responses, coded by the researcher. Agreement reduced, becoming weak, when comparisons were made between the observer's coded notes and participants' self-reported reactions, coded by the researcher.

Taken at face value, the results indicated that the person's report of their internal experience and what was observable to a third-party were different. Additionally, it appeared that the observer's opinions and what their notes indicated, when marked by a blind rater, were in reasonable agreement. Finally, that there was limited agreement between participants' reports of their internal feelings and the coding of the observer's notes by the researcher. The latter indicated that there are perhaps too many steps in the process which could influence the accuracy of the ratings.

The discrepancy between what was observed and self-reported could in part be explained by the difficulty that many autistic people and people with learning disability experience when expressing their feelings. For autistic people, emotion perception difficulties are commonplace (Harms et al., 2010), which could influence both how they report their internal experiences, and how they express these feelings outwardly.

Notably, many people diagnosed with autism experience co-occurring alexithymia (see: Poquérusse, Pastore, Dellantonio, & Esposito, 2018), which can be characterised as the lack of ability to express mood or emotion both in oneself and others (Lesser, 1981). The impact of this is that people who experience both are unlikely to be able to report accurately on their internal emotional and sensory state (Guilbaud, 2007; Poquérusse et al., 2018). Further, this may influence the recognition and interpretation of emotional state by an observer. Alexithymia is also known to influence the outward appearance of emotion, which would make it difficult for an observer to interpret the emotional state that they were viewing (Markram & Markram, 2010; Trevisan, Bowering, & Birmingham, 2016). This could in part explain previous research that indicates non-autistic people perform worse when asked to identify the

emotions of autistic people, compared to those of non-autistic people (Brewer et al., 2016; Sheppard, Pillai, Wong, Ropar, & Mitchell, 2016). It has even been argued that these differences are due to alexithymia rather than autism (Trevisan et al., 2016). However, other studies have failed to find differences between recognition of autistic and non-autistic expressions of emotion, by non-autistic raters (Faso, Sasson, & Pinkham, 2014). At present, these differences are not fully understood.

Many studies indicate that people with learning disability also experience difficulties with recognising emotion (Scotland et al., 2016), although some research suggests that self-reporting of internal emotions can be quite accurate for some people with learning disabilities, but not all (Lindsay, Neilson, & Lawrenson, 1997). Many people with learning disability experience communication difficulties that would make it very difficult to accurately report on their internal emotional states (Bradshaw, 2001). Expressions of emotions are also potentially more subtle in people with learning disability, which may make them difficult to recognise from an observer's point of view (Arthur, 1999). This is especially important if the observer does not know the person being observed well and had not met them prior to the study. Overall, the literature in relation to production and recognition of emotions by both autistic people and people with learning disability, as well as the frequent barriers of recognising and reporting upon internal states, may to some degree explain discrepancies between what was reported and observed.

The results must be viewed in light of the study's limitations. Firstly, the question asked to participants was, 'What was your reaction to this video?' and participants answered this by stating both what they physically did in response to the video (e.g. responded that they laughed) and at other times, responded with how they felt (e.g. that they were amused). The different ways in which people responded may, to some degree, also explain the differences between what was observed and self-reported, as the observer would have no way of knowing what a person's inner state was if there was no change in outward appearance. Secondly, the observer had little experience working with children or autistic people, and had limited instruction on how to score the responses; both of which were likely to have influenced the results. The observer was only instructed to write down if the person reacted in a way that indicated they had mimicked the video and to record what they saw. This was done to keep the study as straightforward as possible, but more robust instructions about scoring are required to help increase reliability. A further

barrier was that the observer was having to carry out a number of tasks at once, for example, observing the participant and noting their response. In future, participants' responses could be recorded via video and analysed at a later point. This would enable multiple raters to score the reactions and provide the ability to re-watch reactions for more nuanced and accurate scoring of the data. Lastly, a general consideration for the study was that the sample collected was quite small. This was due to administrative delays and school holidays placing a limit on the time available for data collection.

Overall, while there were limitations with this pilot study, a great deal was learned in terms of the practicalities of developing and using new behavioural screening tools.

5.2.7 **Studies 2 – 5: Interrater reliability**

5.2.7.1 **Brief overview**

It is important that screening tools and the way that they are scored is standardised. Glascoe (2005) argues that the administration of screening tools should require a minimal amount of training, that scoring and marking procedures should be simple, and that they have good interrater reliability. This section covers the interrater reliability of the mark scheme used throughout Studies 2 to 5. As discussed previously, people simply self-reported their reaction to the stimuli, this self-report was marked as either indicating mimicry/contagion or not, using a simple mark scheme (see Table 41). In this study, a proportion of the responses were second marked by a naïve rater, blind to the study with no experience of autism research (note that this is different to the naïve observer in Study 5). This was intended to indicate whether a simple mark scheme could be standardised for use by a wide range of people.

5.2.7.2 **Procedure and analysis strategy**

People's responses to each stimulus was recorded in their respective studies using the mark scheme previously outlined in 5.2.3 (Table 41). This was first done for the whole sample by the researcher, then at least 25% sample of responses from each study was taken and coded by a second naïve rater. The naïve rater was a PhD student with no prior knowledge of the study or any involvement in autism research. The 25% included in this analysis is in excess of the recommended 10-20% by Neuendorf (2002). Due to the imbalance between autistic and non-autistic participants in Study 3, this was split

between those identifying as autistic and those who did not. In Study 5, a larger percentage (50%) were included, as there were fewer participants here compared to Studies 2 and 3.

Percentage agreement for the coding of each item, the mean, standard deviation, and range of these percentage agreements was then calculated. Lastly, Cohen's Kappa for each study was calculated, as this is more robust than percent agreement alone due to it considering the influence of chance on the sample (Banerjee, Capozzoli, McSweeney, & Sinha, 1999). The value of Kappa was interpreted using guidelines outlined by Landis and Koch (1977).

5.2.7.3 **Results**

Percent agreement and relevant descriptive statistics were calculated (see Table 45). For many of the items there was a high agreement, the mean of each study was 88% or greater and many items showed complete (100%) agreement.

Cohen's Kappa was used to calculate overall interrater agreement for each subsample, and all were significant. The whole subsample of Study 2 showed perfect agreement ($K = .803$, $p < .001$), both autistic and non-autistic subsamples in Study 3 showed perfect agreement (autistic: $K = .815$, $p < .001$; non-autistic: $K = .804$, $p < .001$), and the whole subsample of Study 5 showed perfect agreement ($K = .910$, $p < .001$).

Table 45: Agreement* on a subset of self-report data from Studies 2, 3 and 5, between the researcher and a naïve rater.

Item number	Study 2 Whole sample (*N* = 14)	Study 3 Non-autistic sample (*N* = 20)	Study 3 Autistic sample (*N* = 5)	Study 5 Whole sample (*N* = 7)
Mean *(SD)*	89.29 (14.01)	91.00 (20.12)	88.00 (17.89)	91.42 (7.83)
Range	46.20 – 100	55.00 - 100.00	60.00 – 100.00	85.70 – 100.00
6	92.90	100.00	100.00	100.00
7	78.60			
8	92.90			
9	92.90	100.00	80.00	85.70
11	46.20	55.00	100.00	85.70
12	85.70			
13	100.00			
15	92.90			
17	85.70	100.00	60.00	100.00
18	100.00			
23	92.90			
24	100.00			
25	78.60			
28	100.00			
36	100.00	100.00	100.00	85.70
37	92.90	95.00	50.00	100.00
40	100.00	55.00	100.00	100.00
41	78.60	65.00	80.00	71.40
44	100.00			
45	100.00	100.00	100.00	100.00
46	92.30			
48	78.60			
49	100.00	90.00	80.00	100.00
50	78.60			
51	92.90	85.00	100.00	100.00
61		100.00	100.00	100.00
62		100.00	100.00	100.00
63		94.40	100.00	100.00
64		100.00	100.00	100.00
65		100.00	75.00	100.00

*Note. Agreement is valid percent agreement.

5.2.7.4 **Discussion**

These interrater reliability statistics indicated that it is possible to develop a reliable and consistent mark scheme for self-reported reactions, which can be used by a naïve rater who is given minimal training. This finding differed from that in Study 5, where a discrepancy was found in terms of how the observer scored a response and the way in which their notes about the same observation was scored by an independent rater. This result was in the absence of specific instruction to the observer and a clear marking system.

The present analysis indicated that two raters, with varying levels of involvement in the study, will reach similar conclusions when using a specific mark scheme, even with very little training. This showed how essential a clear and concise mark scheme is for use by those using screening tools (Glascoe, 2005).

In conclusion, the analysis of interrater reliability of scored self-report responses, across these studies signified that a simple mark scheme can produce reliable results. This is important and it means a straightforward, yet specific mark scheme can be applied by raters in a further refined version of a behavioural screening tool.

5.3 General Discussion

Each of the five studies and the accompanying analyses were steps in the development of a screening tool. To briefly summarise the aims and findings of these respective studies, Study 1 (5.2.2) set out to find videos that produce consistent reactions, which it successfully did, identifying a number to be used in subsequent studies. Study 2 (5.2.3) built upon this, using the selected videos alongside four additional videos. From this study, those with strong reactions were identified, meaning a shorter task could be used in subsequent studies. In Study 3 (5.2.4), these videos and a further five yawn videos were shown to participants, although how many yawns a person saw before yawning could not be analysed due to the way the data was collected. In Study 4, machine learning analysis using cforest (5.2.5) provided an important insight; the analysis indicated that it is indeed possible to estimate a person's AQ score from their self-reported responses when watching a video, although the margin of error was quite high. Study 5 (5.2.6) was centred around the trialling of observational analysis of younger people. In this study,

adolescents viewed videos and self-reported their responses, while an observer reported on their reactions. This study found a discrepancy between what was self-reported and what was observed, the implications and potential explanations for this were discussed. Finally, the analysis of interrater reliability (5.2.7) indicated that a simple mark scheme was robust when used to score the self-reported behaviours of participants. When taken together, these studies show that by combining various existing stimuli that in some way relate to empathy, mimicry, and emotional contagion, by asking a straightforward question about responses, and by applying a simple mark scheme, you can estimate a person's AQ score.

5.3.1 **Issues and limitations**

The different studies also highlighted various issues which would need to be addressed in the development of a final behavioural screen. First, the videos were not standardised, as they were pre-existing. While the final videos did produce relatively consistent responses, the differences in picture and audio quality could have influenced the results of any of the studies. For example, some items may have evoked a strong reaction due to simply looking nicer, rather than the content. In addition, the video backgrounds varied between stimuli and studies. These issues can be rectified in future work by including videos with standardised backgrounds that are all filmed to the same high quality.

Secondly, when combining Study 2 and 3 to perform a machine learning analysis, the specifics of the results should be viewed with caution. While the main finding that AQ score can be estimated from responses to videos is encouraging, there are some caveats. These include that the Pain and Laughter videos were the best performing stimuli and the two Disgust videos were the poorest. These should be viewed prudently as participants in these studies had also viewed additional videos, the responses to which were not included in this analysis. For example, those participants in Study 2 would have seen an additional Scare video, while those in Study 3 would have seen additional Yawn videos. This may have influenced the results; one indicator being the low contribution of Yawn videos in terms of significance, despite research that indicates yawning is contagious in non-autistic people, but less so in autistic people (Giganti & Esposito Ziello, 2009; Guggisberg et al., 2010; Helt et al., 2010; Preston & de Waal, 2001; Provine, 1986; Senju et al., 2007). This could be the case because certain responses to videos had been

inhibited, for instance participants may not have been as affected by the Yawn videos due to them having seen multiple during the study. This is analogous to Provine's (1992) finding that participants are less likely to laugh at laughter stimuli after repeated exposure due to fatigue. In future, this can be addressed by only including videos that will be seen by all in the analyses.

Thirdly, many of these studies were based upon self-report. To work best, a behavioural screen should work by observation as it would mean a wider range of people could complete it. Study 5 addressed this and trialled observational analysis with a small group of participants. It was found that there appears to be some difference between what is observed and self-reported. As previously discussed (see 5.2.6), there are various reasons why this might be the case, including the wording of the question, the observer not being able to attend to everything due to doing multiple tasks at once, no robust mark scheme being used, and that both autistic people and people with learning disability may have difficulty in accurately communicating their internal feelings (see: Bradshaw, 2001; Guilbaud, 2007; Poquérusse et al., 2018). To address this, in future the wording of the question should focus on the person's internal state, and the participant's responses to the stimuli should be video recorded. The former would mean the question is more focussed and is not something that could be observed. The latter would free up the researcher to facilitate the study and for them to score participants' reactions at a later point. The ability to review the responses and for multiple raters to score them also facilitates improved accuracy.

In all of the pilot studies, participants were informed as to the purpose of the study which may have potentially affected their reported or observed reactions. For example, Provine (2005) argues that knowing a researcher is interested in contagious yawning reduces the likelihood of a contagious yawn occurring. This is itself a difficult issue to address, as participants will always be informed of the purpose of a study and in clinical practice, those being assessed will always be aware of the purpose of their interaction with a clinician. These are issues which will need to be considered in developing the final screening tool.

There were some issues with the types of videos used in the pilot studies. While they were chosen with the aim of inducing empathic reactions, some depicted an event or action, rather than a facial expression of emotion. Therefore, whilst these videos were able to produce consistent responses and these were able

to predict AQ scores, they are not the types of stimuli which are traditionally associated with mimicry and contagion research. In future, videos should focus more on the mimicry and contagion aspects of the stimuli, rather than actions or events that induce these reactions. This would provide a clearer link between the proposed behavioural screen and the research that it is building upon.

A major omission of this research was the failure to include any itch-based stimuli. The pilot studies focussed on mimicry centred around facial reactions. However, there is a body of literature that focusses on contagious itching, whereby if a person observes scratching they themselves will scratch (e.g. Niemeier & Gieler, 2000; Ward, Burckhardt, & Holle, 2013). The literature on contagious itching will be discussed in the next chapter.

5.3.2 **Conclusion**

In conclusion, the pilot studies presented here indicate that it is indeed possible to estimate AQ scores from reactions to videos. This is very promising, as it opens up the possibility of adapting this idea into a far more accessible screening tool, that is relatively simple to use for both the people delivering it and those taking part. While there are many limitations to the pilot studies, they have been vital to inform the development of the behavioural screen. Specifically, investigating theoretical ideas that had not previously been combined, such as including stimuli that link to mimicry, emotional contagion, contagious yawning, contagious laughter, and depicting animals and humans. The series of studies also provided practical points to consider, which includes the best way to record responses. Overall, these studies represent a first, but significant step toward developing more accessible screening tools.

6.0 Chapter 6: Developing the behavioural screening tool.

ESM: https://osf.io/e3728/ Materials found in folder titled 'Chapter 6'

6.1 Introduction

The previous work presented so far in this thesis indicates that autism screening for people with learning disability can be challenging. The autism screening tools that can be used with this group show questionable validity and reliability, and predominantly rely upon the reports of third parties (Chapter Three). While attempts to adapt the AQ into a tool that is accessible to people with learning disability was somewhat successful (Chapter Two), this self-report format is still not suitable for many people with learning disability. This indicates that a new more user-friendly type of screening tool is required. How best to deliver such a tool was explored through interviews in Chapter Four. In Chapter Five, the idea of a behavioural screening tool was developed through a series of pilot studies that employed a range of stimuli, to which people's responses were coded. This coding was analysed with machine learning using the cforest package in R (Hothorn et al., 2010; R Core Team, 2013) where it was shown that the coded self-reported reactions were able to predict a person's AQ score, albeit with a margin of error. In the current chapter, the idea of a behavioural screening tool has been taken further and focusses on using salient, non-verbal behaviours as the tool's basis. Some literature considered in Chapter Five was particularly pertinent to the current study and as such, it is revisited here.

Empathy can be thought of as having two constituent elements, the emotion based affective component and the perspective taking cognitive component (Cuff, 2016). Some researchers argue that autistic people, when compared to their non-autistic peers, have a reduced capacity for empathy (Baron-Cohen & Wheelwright, 2004; Bird & Viding, 2014; Frith & Happé, 2005; Lawson et al., 2004). Other researchers disagree and instead suggest that empathy may only be reduced in certain situations, that some individuals do not show deficits, or that empathy merely develops later (Fletcher-Watson & Happé, 2019; Tager-Flusberg, 2007). Similarly, in a review of autism and empathy, Song and colleagues (2019) found that when empathy is broken down into multiple domains, a reduction is only found within cognitive empathy and empathic concern (specific responses to suffering), but not affective empathy.

While the specific nature of empathy in autistic people continues to be debated, it does appear that there are subtle differences in empathy and its development in autistic people.

The research into empathy has been extended to work on mimicry and emotional contagion. In a comprehensive review, Hatfield et al. (2014) outline that, when an observer views an emotionally expressive face, their own face will change to more closely resemble and mimic the expression that they are looking at. It appears that this happens irrespective of whether the stimuli is a static image or a dynamic video, and regardless of whether the person was prompted to mimic or simply exposed to the expression (Moody et al., 2007). Related studies into emotional contagion have shown that in both lab-based research (Barsade, 2002; Hennig-Thurau et al., 2006) and real-world interactions (Kramer et al., 2014; Pugh, 2001), not only do people's expressions become congruent with the expression being depicted, but their moods also become more closely aligned. Mimicry, emotional contagion and empathy may be interrelated. Hatfield et al. (2009) proposed that a person first mimics the expression that they see, then feel feedback due to the mimicry, which results in emotional contagion as the person's mood reflects the new affective state. Empirical research that shows the interlinking of these facets is not uncommon. Some researchers show that individuals with lower empathy indeed display lower mimicry (Seibt et al., 2015), while others show that individuals with higher empathy scores are more likely to mimic and experience emotional contagion (Sonnby–Borgström, 2002). This can also reportedly occur when viewing videos of emotional expressions (Rymarczyk et al., 2016).

Research on autism, mimicry and contagion has been carried out. Findings show that when researchers display expressions of emotions to children, those with autism are about half as likely to display a congruent emotion in response, compared to children with either developmental delay or those who are typically developing (Scambler et al., 2007). Similarly, mimicry is reduced in children with autism when they view images of expressive faces, but not when they are directed to mimic. Under these circumstances they copy expressions just as well as control participants (McIntosh et al., 2006). Similar findings extend to autistic adults and when using videos stimuli (Yoshimura et al., 2015). A review of the phenomenon concludes that, while learning disability may influence reduced mimicry, it cannot fully

explain this difference and therefore, this mimicry reduction is likely due to autism (Vanvuchelen et al., 2011).

Significantly less research has been conducted with autistic people in respect of contagion, but related research can provide some insight. In fMRI research, participants were presented with expressions of neutral and fearful body language poses and separately, neutral and pained facial expressions. Autistic adults compared to non-autistic control participants show reduced reactivity to fearful, but not pain, stimuli. According to the authors, these findings indicate reduced emotional contagion, but apparently not universally for all types of emotional display (Hadjikhani et al., 2009, 2014). In all, there is less certainty around differences in emotional contagion, but it is likely mimicry of emotion is reduced in autistic people.

There are a multitude of other behaviours that humans tend to mimic. There appears to be differences in the prevalence of mimicking laughter and yawning, between autistic and non-autistic individuals. Laughter can be considered both a socially performative cue (Provine, 2004; Provine & Emmorey, 2006) and an emotion (Scott et al., 2014). Similar to other emotive expressions, evidence shows that hearing and observing laughter can trigger laughter in observers (e.g. Hatfield et al., 2014; Neves et al., 2018). Again, people who are autistic appear to be less likely to contagiously laugh in this manner compared to non-autistic controls (Helt & Fein, 2016; Reddy et al., 2002). As for contagious yawning, although this is not an emotion, it has been reproduced across numerous studies with both humans and animals (for review, see: Guggisberg et al., 2010). When people see others yawn, they will likely feel the impulse and go onto yawn themselves, this response appears to be consistent from around four or five years of age (Anderson & Meno, 2003; Giganti & Esposito Ziello, 2009). As with other types of mimicry and contagious behaviour, contagious yawning appears to be reduced in autistic people. In children with autism, there is evidence that when they are exposed to yawn video stimuli they are less likely to yawn than IQ-matched control participants (Senju et al., 2007) and are less likely to when yawns are performed in person (Helt et al., 2010). These results appear to be unaffected by autism severity (Giganti & Esposito Ziello, 2009). Together, this research shows that mimicry is not strictly limited to facial

expressions of emotions but extends to other behavioural features and is seemingly unaffected by autism severity.

One type of stimuli not covered in the previous chapter is itching. According to Niemeier and Gieler (2000), itching is the sensation which is associated with the wish and impulse to scratch. In their research, Niemeier and Gieler delivered an itch-inducing lecture for a television show. The participants were filmed for the show and for the purpose of coding whether or not they itched while attending the lecture. In said lecture, the researchers showed images and discussed fleas, mites, scratch marks on the skin, allergies, amongst other potentially itch evoking imagery. Afterward, participants were shown slides containing relaxing stimuli to induce a sense of well-being. The results from questionnaires and reviewing the footage of participants showed that being exposed to itch-related stimuli, led to participants scratching more and feeling itchier than when viewing the relaxing stimuli. While this is not mimicry per se, it does show that itching can be induced in participants. Papoiu, Wang, Coghill, Chan, and Yosipovitch (2011) researched the mimicry of itching behaviour more directly. In their work, participants viewed five-minute video clips of people scratching or sitting idle and were monitored by the research team. Participants were found to scratch more often when viewing the itch compared to the idle stimuli. This effect was apparently amplified in participants who had pre-existing skin conditions, a finding also present in Niemeier and Gieler's research. Further investigation was conducted by Holle, Warne, Seth, Critchley, and Ward (2012) who replicated the finding that viewing people scratching led to itching. They found that while viewing scratching of all parts of the body led to contagious scratching, itching the upper left arm led to the greatest response.

These findings are mirrored in primate research that show when monkeys watch videos of other monkeys scratching, they will scratch in response; this implies a widespread evolutionary phenomenon (Feneran et al., 2013). Curiously, this does not seem to be the case in rodents (Lu et al., 2019), which do not show any contagious itching behaviours. The authors argue that this could be related to the lack of emotional contagion and affective empathy in rodents.

Contagious itching and its relationship with autism has also been investigated, although less frequently than research into other forms of mimicry. Work by Schineller (2018) investigated yawn, laughter, and

itch stimuli in a group of children with autism. While reduced contagious laughter and yawning was found in the children with autism, contagious itching appeared heightened compared to the control group. Similarly, in an eye gaze study Scheub, Sorenson, and Helt (2020) found that increased AQ scores were associated with a decreased tendency to yawn, but an increased tendency to itch. Sorensen (2017) conducted research that found itching appears to be unaffected by eye contact or empathy, which is at odds with findings from facial mimicry-based research. This is also somewhat at odds with the proposal of Lu et al. (2019) that this itch response is based in affective empathy. Overall, there is very limited research in this area. If itching in response to others scratching is due to empathy, it might be expected that autistic people would be less likely to contagiously itch, though empirical evidence suggests the contrary. Despite this, there still appears to be differences between autistic and non-autistic individuals in contagious itching. It can be argued therefore, that stimuli designed to provoke this type of response may prove a helpful addition to the behavioural screening tool.

Chapter Five showed that participants' AQ scores could be estimated from the prevalence of mimicking behaviours and emotional contagion. This is in line with a body of previous research that shows differences between autistic and non-autistic individuals in terms of these behaviours (e.g. Helt & Fein, 2016; Schineller, 2018; Senju et al., 2007; Yoshimura et al., 2015). While the data in Chapter Five were mostly based upon self-report responses, the present study measured both the self-reported responses of participants as well as their non-verbal behaviour in response to stimuli. This was observed directly by video recording their reactions. Both the recorded and self-reported responses were coded in terms of whether the response indicated mimicry or emotional contagion. It was hoped that the additional focus on non-verbal responses would extend the accessibility of the proposed screening tool for autism, allowing it to be used with those who have limited or no verbal communication. For the purposes of the present study, there was a focus on recruiting individuals with a diagnosis of autism and/or learning disability, in order to evaluate the accuracy of the screening tool at differentiating between people in terms of their diagnostic status. Where possible, AQ responses were also gathered.

Drawing on previous research and the results of the studies outlined in previous chapters, bespoke stimuli for inclusion in the screening tool were created for this study. This was to ensure that the quality,

style and presentation of the videos were more consistent. Before moving onto the main study, an overview of the development of the stimuli is provided here. The development included interviews in which participants suggested ideas for stimuli, that may lead to differences in mimicry and contagion between autistic and non-autistic people.

6.1 Developing behavioural screen stimuli

In order to ascertain a baseline of the participant's emotion, a Neutral video that was devoid of any emotional content was created. Then, a standardised array of videos containing facial displays of emotion was created that included Happy, Sad, Angry, Disgust, and Fear. Finally, videos that depicted common mimicry inducing behaviours such as Pain (which was found to be the best predictor of AQ score in Chapter Five), Yawning, Laughing, and Itching, were made.

To account for the possibility that other potentially differentiating types of stimuli existed, that are not reflected in past research literature, individuals were approached who had experience or expertise relevant to autism (see Chapter Four). These people were asked about potential ideas for stimuli that may lead to differing reactions. The proposed ideas were subsequently assessed in terms of feasibility.

6.1.1 Further stimuli ideas

At the end of the interviews reported and discussed in Chapter Four, participants were asked for ideas about stimuli that could be included in the behavioural screening tool. The participants made various suggestions, a number of which were implemented in the current study. Table 46 provides a quote relevant to each idea, a short commentary on the idea and an overview of how it was then implemented into the stimuli set. The ideas taken from these interviews included showing participants non-verbal instructions, stimuli that relate to the different sensory experiences of autistic people, content that may show differences in empathy, and different reactions to pain. A number of other ideas were suggested but could not be implemented because of practical issues, such as not fitting into the format of a short video clip. For further detail, see the ESM (Chapter 6 / Appendix Chapter 6).

Table 46: Extracts from interviews in Chapter Four on the discussion of creating items for a behavioural screening tool.

Idea & example quote*	Commentary	Implementation
Following a non-verbal instruction		
Perhaps people following some non-verbal instruction (C30)	During the interviews the idea of giving a non-verbal instruction was proposed, in which autistic and non-autistic participants may respond differently to each other.	To implement this, a video was created in which a person points at the camera and then at their own head, twice. Participants were observed to see whether they point at or attend to their own head in return.
Sensory differences		
Sensory things … I'm envisaging the same video which is just a sort of generic video, but at three different volume points … and you could watch their reaction as the volume goes up … a calm street, a busy street, and how they react to that (T2) *You could have noises [my son's] quite sensitive to noises we've got a friend that lives round the corner whose sensitive to noise … You could have something [repeated taps] and someone is talking you could it on a different speaker or something in the corner you could be the person who turns it up and when someone reacts to it they would have quite a high sensitivity to noise … a distraction yeah (P25)*	The ideas presented by participants here relate to different sensory experiences of autistic people. These could be implemented in a video form but would have to be done in a careful and considerate manner. One idea proposed was showing different streets, one busy and one quiet. Another idea was to use a video of a person, with a tapping noise in the background. Interviewees proposed that there may be observable differences between autistic and non-autistic participants when they attended this type of stimuli.	These ideas were implemented in two different types stimuli. Firstly, two video clips that contrasted a busy with a quiet street and secondly, a video that shows a neutral face with a tapping noise in the background. Participants were scored in terms of whether they appear to be more engaged or distracted by either of the types of video. Comparisons were relative to the neutral or the quiet videos.
Empathy: Seeing someone get bullied or fall down		
I mean in a really basic manner, some child calling another child erm like 'Shrek' and they're really quite obese and not really good looking and see what the child you're assessing does (T2) *Two people maybe somebody pushing the other person or something like that, whether that sort of made, gave a response cos I think even a young child generally would kind of … Look a bit *gasp* they shouldn't have done that you know and that's not always, that's not always the same response you'd get (P19)*	Many participants commented on differences between how autistic and non-autistic people react in situations where another person has been insulted or harmed in some way. With the literature demonstrating mixed findings in this area (Bird & Viding, 2014; Tager-Flusberg, 2007), it was included on an exploratory basis, as multiple participants proposed the idea.	In order to ensure the stimuli are accessible to non-verbal participants, the suggested concept was operationalised by creating a video of one person pushing another, with no background, contextual information. The videos were gender counter-balanced, in one a male pushes a female and in the other a female pushes a male. Different actors were used in each stimuli. Responses to the videos were observed and coded in terms whether the participant reacted empathetically toward the stimuli.

Pain		
[if they saw someone] [fall] down or hurt their finger, something that you would go 'aww' … the empathy thing that I'm thinking about really so they can sort of show some emotion to feel with that person with what … happened (T9) *I've seen much more variation [of empathy] among autistic people I know than I have in neurotypical population but at both ends of the extreme that people are completely incapacitated by others in pain (A20)*	Some participants discussed how autistic, compared to non-autistic, people may react differently to another person's pain. There was not a consensus as to whether their reaction would be more or less extreme. Pain-centric stimuli had featured in Chapter Five and was found to be predictive of AQ score.	Pain stimuli was already planned to be included in the behavioural screen, but the differing suggestions about the direction of reaction in autistic people were interesting to note. Reactions were coded in terms of whether the participant winces, empathises or shows an apparent pain response to the stimuli.
Adrenaline		
I'm thinking as well like for the boys in particular, they love speed and it's all almost like that adrenaline fix but from a safe distance so like a car doing donuts or something … I think it's sort of, like sort of sit up and take notice for something like that … it's exciting and it's thrilling like you know for instance, you know there's like a camera on one of these thrill rides like a rollercoaster or something like that (T9) *Things with danger and stuff like that like that guy who jumps off the bungee jump without a rope that's dangerous innit see what their reaction is (P25)*	Both of these ideas related to adrenaline and how autistic and non-autistic participants may react differently to this type of stimuli. As with the pain stimuli, it was unclear how autistic and non-autistic people might differ in their reactions.	As it was difficult for the research team to film an adrenaline inducing scene in a controlled environment, the idea was instead implemented by using pre-existing videos of rock climbers falling. Reactions were coded in terms of whether the person showed surprise, shock, or appeared more energetic and attentive, compared with responses to the neutral videos.

Note. Further quotes are available in the ESM related to many of the ideas implemented here (Chapter 6 / Appendix Chapter 6).

6.1.2 **Stimuli creation**

6.1.2.1 **Final stimuli and equipment**

The emotion, behavioural mimicry, and contagion stimuli ideas outlined in 6.1, and the further ideas from participants discussed in 6.1.1 were compiled in Table 47, alongside descriptions of what the stimuli included.

In order to create a task that could check whether participants were attending to the stimuli, some of the actors featured in the stimuli were asked to wear a red t-shirt. Participants were then asked to say when they saw someone wearing red after viewing each video clip.

A recording space was set up at the university, in a quiet area with good lighting. A white sheet was used to standardise the background of these recordings. A Sony 4K video camera (1920x1080) with an external microphone was used to film the stimuli. It was tripod mounted to centre it directly toward the actor's face. The researcher did not feature in any of the videos due to the potential influence of familiarity on mimicry and emotional contagion (Helt & Fein, 2016).

Table 47: Stimuli to be created as part of the stimuli creation project.

Video	Comment
Neutral	Control stimuli, where the actor gazes toward the camera showing no emotion. The aim was to provide a baseline reaction of participants' neutral expressions, against which the reactions to the other stimuli could be compared for the purpose of coding.
Happy	Emotion stimuli in which an actor smiles and appears pleased.
Sad	Emotion stimuli in which the actor looks sad, upset, and down.
Angry	Emotion stimuli where the actor looks angry and annoyed.
Disgust	Emotion stimuli where the actor looks disgusted.
Fear	Emotion stimuli where the actor appears scared or frightened by something out of the camera's view.
Pain	Emotion and behavioural stimuli where the actor expresses pain. This effect was created by delivering a small electric shock to the actor's hand.
Itch	Behavioural stimuli where the person scratches and looks uncomfortable. Actors were instructed to start at the upper left arm as this appears to be the most highly contagious area (Holle et al., 2012)
Laugh	Emotion and behavioural stimuli where the actor laughs. This was created by first having the actors 'fake' laughter which subsequently triggered 'real' and convincing laughter.
Yawn	Behavioural stimuli where actors were instructed to yawn. If they struggled to do so, they were asked to 'fake' yawn and the researcher also 'fake' yawned to provoke a behavioural response on camera.
Adrenaline	Interview study idea, which was not filmed by the researcher. Instead edited footage of rock climbers falling was provided by Hot Aches Productions, an adventure filmography company[2].
Point	Interview study idea, based on non-verbal instruction where the actors were asked to point at the camera and then point at their head, twice.
Push	Interview study idea based on empathy differences. One actor was asked to walk up to another actor, push them and walk away. The actor receiving the push was instructed to look upset and annoyed at what had happened.
Distract	Interview study idea based on sensory differences. The actor was asked to show a neutral face while the researcher taps out of sight of the camera.
Quiet	Interview study control stimulus idea. This was used as a control-stimuli to provide an indication to the coder of the participants' responses to a quiet street, as compared to the stimuli depicting a crowded street (see below). Footage was taken on a quiet street.
Crowd	Interview study idea based on sensory differences. This was filmed on a busy shopping street near the university.

6.1.2.2 **Procedure**

This research gained ethical approval from Northumbria University's Health and Life Sciences ethics committee via the online ethics platform (reference number: 13775). In this study, actors for the stimuli were sought through word of mouth. Potential actors were briefed and made aware that they would be asked to perform emotive expressions and actions toward a camera, which other people may later view and report reactions to. If they were interested, a time was arranged for them to visit the recording booth

[2] https://www.hotaches.com/

at the university. Actors were either asked to wear a plain coloured top (e.g. black, white, grey) or a red top during filming. Upon arrival full information about the study's purpose, how the collected data would be used and how to withdraw from the study, was provided to the actors before they gave written consent to take part. In addition, they were advised that should they subsequently wish to withdraw their footage, it would be deleted on request.

Actors were told what types of videos were being created and could choose to not film any particular one (e.g. the Pain stimuli). Some were filmed on more than one occasion in order to refine the stimuli. The complete list of emotions and actions can be seen in Table 47, alongside descriptions used by the researcher to prompt actors. It was up to the actors what order they would like to film in. For each video, actors were asked to start by assuming a neutral face before performing the intended emotion or action. Each video clip lasted at least 10 seconds. Multiple videos were taken for each type of stimuli and were reviewed by the researcher and actor during the filming sessions, in order to achieve the best stimuli.

The Adrenaline, Quiet, and Crowd stimuli were filmed on location (Quiet/Crowd) or were provided by a third-party (Adrenaline), see Table 47 for details. An example of one final stimulus video is shown in Figure 6.

Figure 6: Example stimuli, Fear (male).

6.1.2.3 **Participants**

Sixteen volunteer actors took part in the development of the stimuli, 7 male and 9 female. No other biographical details were collected. Actors were required to be 18 years old or above. No other exclusion criteria were used.

6.1.2.4 **Behavioural screen stimuli selection**

The selection of the final stimuli for inclusion in the behavioural screen involved several stages, as can be seen in Figure 7. Stage 1 involved excluding several videos due to the actors appearing distracted or because of an issue with background noise. During Stage 2, the remaining videos were watched again by the researcher with the aim of retaining the 2 best quality (e.g. the most realistic and convincing) depictions from each actor. The additional videos for Adrenaline, Crowd, and Quiet were added into the stimuli pool at this point. In Stage 3, the compiled videos were categorised by the researcher and their supervisor, separately, into 'Yes' (retain), 'Maybe' (possibly retain), and 'No' (remove) categories. The researcher and supervisor then met to discuss why each video had been selected and to jointly agree on the retained videos at this stage. A number were selected to be re-filmed, for example because a video had slight background noise or camera glare. In the final stage, actors from the videos which were identified for reshooting were contacted and a second round of filming took place. These new clips were added into the video pool. The researcher and supervisor met again to select the final stimuli set, which was gender balanced between male and female actors in the videos.

The final set of stimuli comprised 32 videos. Two (1 male, 1 female) depicted the following: Happy, Sad, Angry, Disgust, Fear, Pain, Itch, Laugh, Yawn, Adrenaline, Point, Push, and Distract. One video of a Crowd and another of a Quiet street were included. In addition, 4 (2 male, 2 female) depicted Neutral expressions. These videos were edited so that each lasted approximately 10 seconds and had maximum clarity (e.g. brightening the video or lowering the volume).

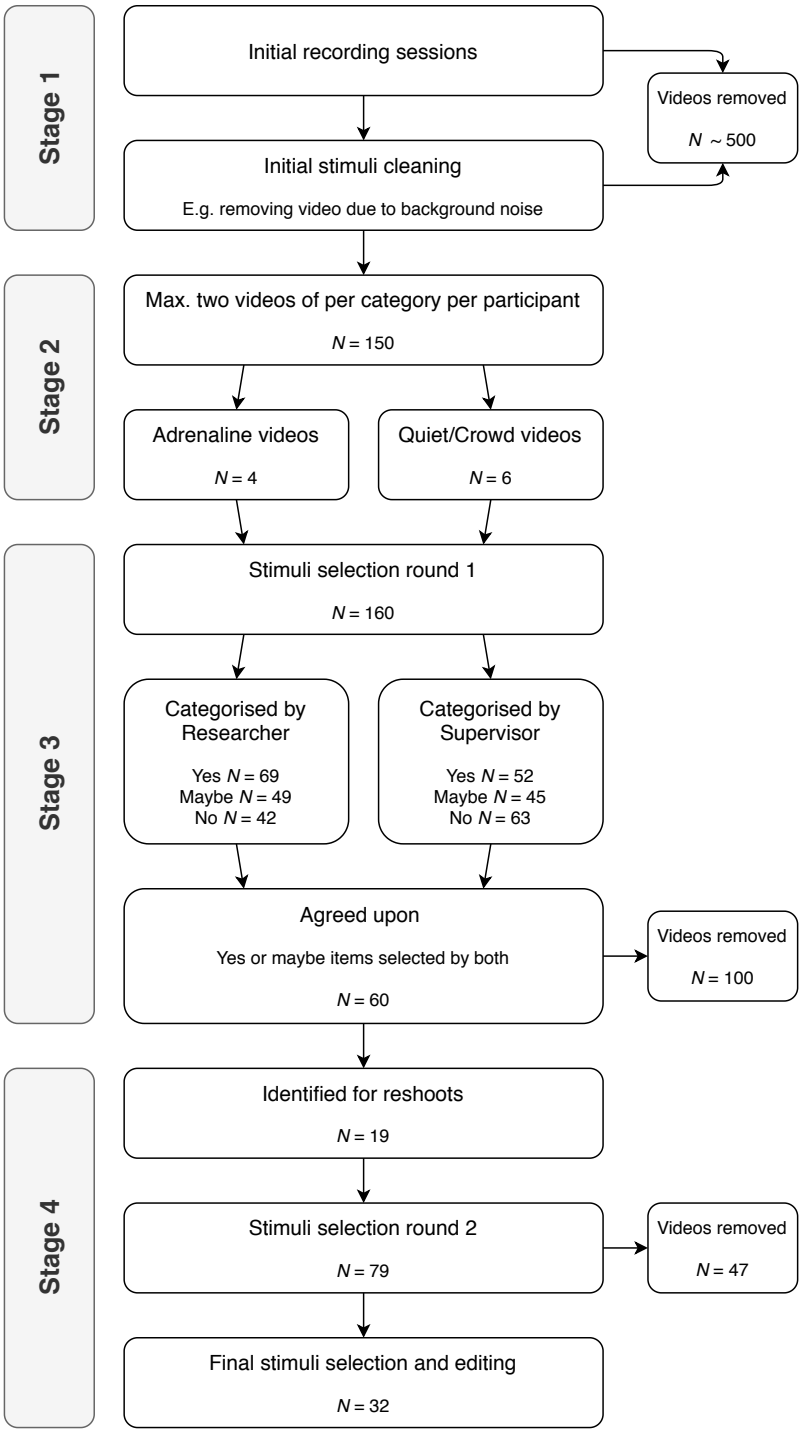


Figure 7: Outline of the stimuli selection process.

6.1.3 **Behavioural screen presentation**

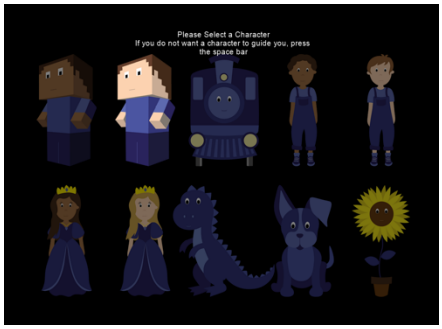
A software application written using the PsychoPy (Peirce et al., 2019) and OpenCV (2020) libraries for Python3 was created to display the stimuli to participants whilst capturing video and self-report data. The programme had an option for all of the written instructions to be narrated to the participants rather than relying on the participant reading themselves, or having the researcher narrate. Each screen of the programme that the participants viewed is shown in Figure 8 (number in square brackets throughout this section refers to the panel of this figure). To make the testing experience more appealing to participants, participants were able to customise some of the visual aspects of the experience. This included choosing a background colour (black, white, lilac, or light yellow) [Fig8, 1], as well as selecting the type of character (e.g. train, princess, dog) [Fig8, 2]; the participant could also choose not to have a character. If a character was chosen, the participant could assign a colour to this avatar (red, blue, or green) [Fig8, 3]. An instruction screen was shown to participants to prepare them for watching the videos [Fig8, 4]. The video presentation segment then began, where all 32 videos were presented in a random order.

Before each video played, a fixation cross was shown in the centre of the screen [Fig8, 5]. The participants pressed the space bar when they were ready to watch the video. Once the space bar was pressed, the video played and simultaneously, the webcam would begin to record the participant [Fig8, 6]. Participants watched the video followed by a waiting screen in which a fixation dot rotated for 5 seconds [Fig8, 7]. This was to allow participants some time before answering questions. The participants were asked, ‘How did you feel when watching the video?’ [Fig8, 8] and, ‘Was there a person in the video wearing red?’ [Fig8, 9]. The former related to emotional contagion of the stimuli, while the latter was used as an attention task to check whether participants were attending the stimuli. For each question the researcher typed a one-word response on the participant’s behalf. After the second response was typed, the webcam stopped recording. This process [Fig8, 5-10] repeated until all 32 videos had been viewed.

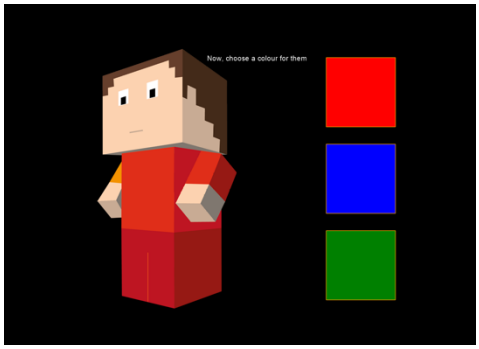
For each participant, the programme produced 32 webcam video clips of participants viewing the video stimuli (without sound), and a tabulated text file that logged the 2 written responses per stimulus. The method of coding these responses is discussed in the next section.

1

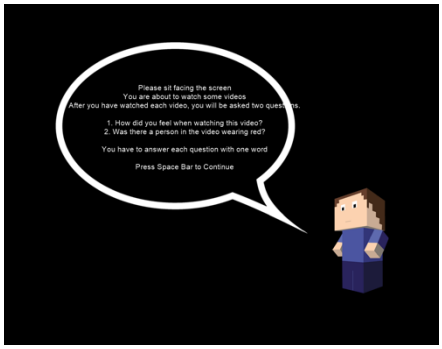
2



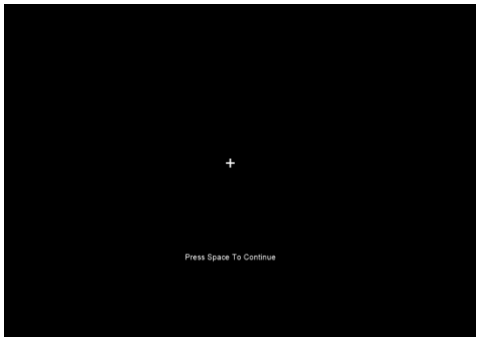
3



4

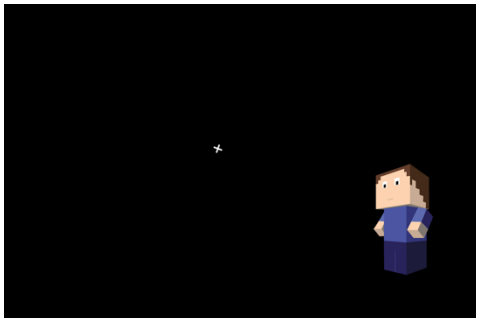


5

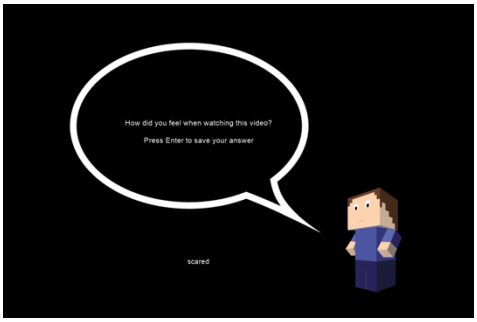


6 Video stimuli displayed full screen

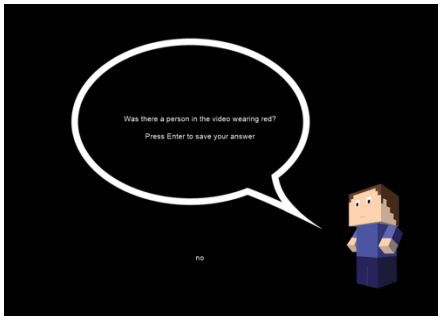
7



8



9



10

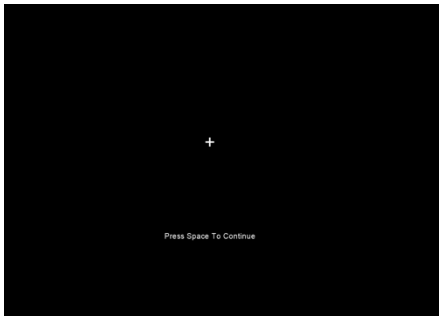


Figure 8: Steps in the behavioural screening tool. Panels 5-10 repeat 32 times, the video displayed in panel 6 changes each time.

6.2 Behavioural screen study

As discussed in 6.1, emotional contagion and mimicry of both facial expressions and behaviours reportedly occur differently in autistic and non-autistic people. In Chapter Five, it was shown that it was possible to predict a person's AQ score based upon their self-reported reactions to video stimuli. The current study builds on published research and the work outlined in previous chapters to further refine the process, and to develop a behavioural screening tool. Foremost, this project aimed to explore whether such a tool can have acceptable psychometric properties in terms of reliability and validity, when used to ascertain whether someone is likely to be autistic. As the final tool aims to have utility with people with learning disability, participants from this group were included in this study both as part of the larger group and also in a subset of analyses.

The current project did not include adults who are non-verbal due to difficulties with gaining consent. However, if it was found to be effective as a screening tool based on behavioural responses alone, this suggests that it may have utility for people with limited or no verbal communication. As the screen's future purpose is to be accessible to people who are non-verbal, the scores created to be assessed via ROC curve analysis and hierarchical regression will always include an observed score. The decision was taken as participants would always be able to provide an observable response to stimuli but may not always be able to articulate a verbal response. This means a verbal response alone is not useful, whereas an observable response and a combination of the two are more likely to be useful, and better able to be collected. This was notably different to Chapter Five which for the most part relied upon self-reported reactions to stimuli.

It was hypothesised that differences exist between autistic and non-autistic individuals in their self-reported, observed, and combined reactions to the stimuli included in the behavioural screening tool; although no hypotheses were made about the exact nature of any difference. This was for two reasons. First, while the existing literature generally indicates that autistic people will likely react in a less congruent manner than non-autistic individuals (McIntosh et al., 2006; Yoshimura et al., 2015), for some types of stimuli such as itching, the opposite seems to be the case (Schineller, 2018). Second, some of the stimuli were based on the ideas provided in interviews with those with experience and expertise in

autism. For these, there was no clear research base to guide hypotheses about what direction a difference, if any, would take.

As a result, the approach taken used logistic regression to explore the discriminating ability and direction of each stimuli. Stimuli were included or excluded from the final behavioural screen according to the outcome of this analysis. The related hypothesis was that the behavioural responses, as indicated by a total score on the behavioural screen, would be able to predict the person's AQ score and diagnostic status.

6.3 Method

6.3.1 Ethical Approval and Design

Ethical approval for this research was provided by Northumbria University's Health and Life Sciences ethics committee via the online ethics platform (reference number: 15443). The study was a between groups observational design, between autistic and non-autistic people, with and without learning disability.

6.3.2 Recruitment

Participant recruitment was targeted at autistic people, people with learning disability, and comparable controls, matched on age and gender. This targeted recruitment can be broken down into the following subgroups: autistic people, autistic people with co-occurring learning disability, comparable controls with learning disability without autism, people who are potentially autistic but not diagnosed (e.g. have been referred for assessment), and comparable controls to the autistic and potentially autistic groups. Both children and adults were recruited. The exclusion criteria included children under 5 years old, people aged 18 years or above who were unable to consent for themselves, adults who were non-verbal, and people with a visual impairment that prevented them from watching the video stimuli.

Recruitment was carried out via multiple pathways including word of mouth, through email and through social media (particularly on webpages for parents of children with Special Education Needs). Recruitment also took place through external organisations such as charities and support networks for autistic people, people with learning disability, and their parents. Several external organisations such as

schools supported recruitment for this project. In such cases, full information about the study was provided at a meeting with the appropriate person (e.g. headteacher) and consent was obtained to recruit through the organisation, using email or letter, according to preference.

In all cases, potential participants were provided with information about the purpose of the study, the nature of the stimuli and what they would be asked to do. All participants were given the chance to ask questions either in person, by email, or on the phone. The latter 2 options were provided for any parents/carers who were providing consent on behalf of someone under 18 years old, but who would not be meeting the researcher in person. Written informed consent was obtained for all participants, either provided for themselves or by parents/guardians for those who were aged under 18 years. Those who consented, undertook the study at their home, the organisation they were recruited through, or the researcher's university, according to preference.

Once the age and gender of participants in the experimental groups (autistic, autistic with learning disability, potentially autistic) were known, the recruitment of the control group participants (non-autistic, non-autistic with learning disability) commenced. This followed the same process as for the experimental group; however, it was more targeted as information about the required age range and gender of participants was given, to allow closer matching between the 2 groups.

6.3.3 **Materials**

6.3.3.1 **The Autism-Spectrum Quotient (Baron-Cohen et al., 2001)**

Multiple versions of the AQ were implemented in this study, as outlined below. All of these were discussed in detail in Chapter Two and only a brief overview is provided here.

Self-Report AQ

Participants of all ages were asked to complete a short-form AQ (Allison et al., 2012) about themselves. This AQ was marked using the dichotomous scoring method first implemented by Baron-Cohen et al. (2001) which gives a possible range of 0 to 10. Higher scores indicate higher autistic traits and accordingly, a higher chance of the person completing it being autistic.

Third-party/Informant-rater AQs

Where possible, a third-party such as a parent, carer, or friend was asked to complete an appropriate version of the AQ about the participant in the study (Auyeung et al., 2008; Baron-Cohen et al., 2006).

Adult AQ

In the development of the original self-report AQ (Baron-Cohen et al., 2001) parents of autistic adults were asked to complete an informant AQ about their adult child. As this version could not be accessed, in this study the original AQ was adapted by changing the statements from being about the person completing it, to being about the person they were completing it in relation to. For instance, by changing 'I' to 's/he' or 'him/her'.

Adolescent and child AQs

When participants were under 18 years old, parents were provided with either the long-form adolescent (Baron-Cohen et al., 2006) or child AQ (Auyeung et al., 2008), depending on the age of their child. Both contained 50 items, of which half were reverse scored.

All informant variants of the AQ were scored using the dichotomous marking scheme (Baron-Cohen et al., 2001), whereby scores were between 0 and 50. Higher scores indicated higher autistic traits and also a higher chance of the person being autistic.

6.3.3.2 **Behavioural Screen**

As outlined in 6.1.2, the behavioural screening tool comprises 32 videos. Participants are presented with these stimuli and are video-recorded viewing each of these video stimuli, while also self-reporting how they feel about each. In addition, as a measure of attention, participants are asked to state whether the person in the video was wearing red or not.

The video recorded responses, the self-reported reaction, and the attention task were all marked dichotomously as 0 or 1. The process for the coding of each is outlined in the sections below.

Coding: Observed Responses

The video recordings of each participant watching each stimuli [response videos] were coded by the researcher upon viewing each video clip. This was facilitated by a Python scoring application which showed the researcher each video clip, clarified which stimulus the participant was watching, and prompted the researcher to score each item. The programme outputted a tabulated text file for each participant, ready for analysis.

The researcher initially watched the participants' responses to the Neutral and Quiet clips to obtain a baseline of how each participant looked when they were not displaying any mimicry. Following this, the researcher viewed the response videos in the order that the participant viewed each stimulus. Each response video was labelled with the type of stimuli that the person was viewing (e.g. Happy). In the response video viewer, a timer of 15 seconds was displayed which indicated when the video and waiting screen have both finished. During these 15 seconds, the participant was coded in terms of either displaying some form of mimicry or congruent response, or neither. Exceptions to the 15 second timespan were the yawning and itching stimuli, as responses to these stimuli appeared to be more delayed. In these instances, the whole response video file was viewed. The researcher was able to replay the response videos as many times as required in order to be certain about the response.

The researcher coded the responses according to whether the participant displayed mimicry or contagion (coded as 1), or neither (coded as 0). The scoring criteria for mimicry and contagion are outlined in Table 48.

Table 48: Indicators of mimicry and contagion for each video stimulus in the behavioural screening tool.

	Response that indicates mimicry or contagion	
Stimuli type	Video	Self-reported
Itch	Itching, wriggling uncomfortably	Itchy, scratchy, unclean
Laugh	Laughing, signs of amusement, smiling,	Amused, humoured, happy
Yawn	Yawning, stifling a yawn	Yawn, tired, sleepy
Pain	Wincing	Hurting, uncomfortable, empathy
Angry	Frowning, looking angry	Annoyed, frustrated, mad
Disgust	Showing disgust	Horrible, 'ew', nasty, gross
Fear	Looking scared, eyes wide open	Scared, terrified, afraid
Happy	Smiling	Good, joyful, nice, cheerful
Sad	Frowning, unhappy face	Down, upset, concerned
Adrenaline	Looking energised, 'active', eyes wide open, visual suspense	Wow, scary, tense
Distract	Looking distracted, uncomfortable	Irritated, agitated, annoying
Point	Person points at themselves	Point, pointy
Push	Looking shocked, unhappy	Angry, sad, concerned
Crowd	Looking distracted, 'active'	Distracted, uneasy, flustered

Responses or reactions that indicate mimicry, score = 1. Any other, or no behaviours or reactions, score = 0.

Coding: Self-Report Responses

Participants provided a one-word response about how they felt while watching each video. This response was marked within the same programme that facilitated coding of the observed responses. This function took each word and checked it against 2 lists of responses, one that indicated either mimicry or contagion and one that did not. Responses that indicated mimicry or contagion were mostly synonyms of the emotion being portrayed, or a reaction that indicated the participant was empathising with the actor in the video (e.g. 'scared' when viewing the Fear stimuli). If a response was on neither of these lists, the programme highlighted the response. The researcher then added it to the appropriate list and continued to the next response. These lists are available in the ESM (Chapter 6 / Analysis / Data & Code / Word Responses).

Coding: Combined

Combined coding of the self-reported and behavioural response was carried out. Here, a score of 1 was awarded to participants who had scored 1 on either the self-report or observed responses; scores of 0

were given to participants who scored 0 on both. It should be noted that if the participant had missing data in either category, a value of 'missing' was entered.

Coding: Attention Task

Similar to the self-report task, the programme also coded responses to the attention task, i.e. the question of whether someone in the video was wearing red. This programme read each participant's response and scored it as correct (1) if the person's response correctly indicated that someone was or was not wearing red in the video clip, or incorrect (0) in all other cases. During data collection, it became clear that this task did not relate to the person's attention, as participants who were clearly attending the stimuli responded incorrectly. This may have been because they were too focussed on other aspects of the video, while other participants appeared to be answering correctly due to chance. As such, mean scores will be reported but no further use of the attention task will be made.

It should be noted that, while all participants provided biographical information and completed at least some of the behavioural screen, some did not complete all measures included in this study or only partially completed some. The behavioural screen also ceased recording for some participants due to technical difficulties ($N = 14$).

6.3.1 **Procedure**

People who were interested in participating in the study were provided with detailed information about it. This included why the research was being carried out, the background to the research, what the potential outcomes may be, how any data (including video footage) they provided would be marked and stored, and contact information so that they could ask any questions. This information was always provided to the individuals taking part and for those who were under 18 years old, it was also provided to their parents/carers.

Arrangements were made for participants to complete the study at a place of their choosing, such as their home, at the premises of the organisation through which they were recruited (e.g. school; the researcher's university). At this meeting, the participant was reminded about the study information. If participants were under 18 years old, written consent from a parent or guardian was obtained in advance. All of those

aged 18 or over, provided written consent for themselves at the meeting with the researcher. If any informants (e.g. parents, carers, other family members) were involved, they were also provided with full information about the study and gave written consent for their input.

For all participants biographical information was gathered including age, gender, ethnicity, and relevant diagnoses (e.g. autism, learning disability, other conditions). This was provided by parents or guardians for those aged under 18. The AQ was completed by parents, guardians, or other third-party informants, either at the meeting or in advance if they were not going to be present during the completion of the behavioural screen. In the latter case, the participant brought the completed measures with them. Participants then completed the Behavioural Screen. They were given the opportunity to take breaks between each stimulus and it was made clear that they were able to stop at any time, for any reason. If participants were happy to do so, they were asked to complete a self-report version of the AQ. Afterwards, participants were debriefed, given the opportunity to ask any questions and again provided with the researcher's contact information. Participants could also opt to receive follow-up information about the study via email.

6.3.2 **Participant summary**

It is worth noting at this point, that the recruitment and testing had to be stopped due to Covid-19 and the subsequent lockdown in the UK. At this time the majority of autistic participants had been recruited and although many had been tested, there were a number who had testing planned for the weeks and months ahead. The impact of Covid-19 restrictions mostly affected recruitment of the control groups, both people with and without learning disability. This meant that while this study included many autistic people, the planned comparison with those without autism was underpowered. It is anticipated that in future when testing can take place again, further participants will be recruited for this study. In the meantime, this thesis was written based on the analysis of the data available at the point that Covid-19 restrictions were imposed.

When recruitment and testing ceased, a total of 158 participants had taken part in the research. All participants were from the UK, 1 participant was white European, and the remaining were white British. Demographic information, broken down by group, is shown in Table 49.

Table 49: Demographic information for participants, broken down by group.

Group		Age M *(SD)*	Gender*	Comments
Whole sample	All *N* = 158	21.27 (15.74)	N_M = 100 (63.3%) N_F = 56 (35.4%) N_O = 2 (1.3%)	Other cond. = 84 (53.2%)
No Learning Disability	Autistic *N* = 74	18.91 (12.56)	N_M = 51 (68.9%) N_F = 21 (28.4%) N_O = 2 (2.7%)	Other cond. = 44 (59.5%)
	Potentially autistic *N* = 16	27.44 (19.07)	N_M = 9 (56.3%) N_F = 7 (43.8%)	Other cond. = 7 (43.8%)
	Non-autistic *N* = 22	27.36 (17.65)	N_M = 8 (36.4%) N_F = = 14 (63.6%)	Other cond. = 7 (31.8%)
Learning Disability	Autistic *N* = 22	14.18 (8.66)	N_M = 20 (90.9%) N_F = 2 (9.1%)	<u>Level of Learning Dis.</u> Mild = 6 (27.3%) Mod. = 9 (40.9%) Severe = 3 (13.6%) Unknown = 4 (18.2%) Other cond. = 17 (77.3%)
	Non-autistic *N* = 19	29.89 (21.83)	N_M = 10 (52.6%) N_F = 9 (47.4%)	<u>Level of Learning Dis.</u> Mild = 6 (31.6%) Mod. = 4 (21.1%) Severe = 5 (26.3%) Unknown = 4 (21.1%) Other cond. = 7 (36.8)

*Note. N_M = Male, N_F = Female, N_O = Other

6.3.3 **Analysis plan**

A power analysis for ROC was first performed for the purposes of sample size calculation.

The main analysis utilised binomial logistic regression that aimed to predict whether a person is autistic or not. As such, those who identified themselves as likely autistic (i.e. the potentially autistic group) but had not been diagnosed, were not included in this analysis. To conduct these regressions, first collinearity of the predictor variables was assessed to ensure the data was suitable for regression. This was assessed by creating a correlation matrix of all predictor variables and identifying any 2 that correlated above .8 (Berry, Feldman, & Stanley-Feldman, 1985). Then by calculating Variance Inflation Factor (VIF) and highlighting variables that had values in excess of 5 (James, Witten, Hastie, &

Tibshirani, 2013). Variables were removed if collinearity was problematic, variables with a stronger basis in the literature took precedence over the more exploratory stimuli included in this study.

Three binomial logistic regressions were run on the suitable data. The first on the observed responses, the second on self-reported responses, and the third on the combined score of both. Although the self-reported only total score was not used in subsequent analyses, it was included at this stage as an exploratory analysis to assess if the self-reported responses alone may be able to discriminate for people who are able to complete this part of the screening tool. Regressions were assessed in terms of overall model fit and subsequently, any items that contributed significantly to the model.

If the model was found to be a significant fit, any items found to be significant were retained to contribute to the total score. If the model was not found to be a significant fit, all items would be retained for the total score. Items that contributed to the total score may be reverse scored according to the beta value found within the binomial regression. Total scores were calculated for the observed responses and the combined responses, but only for the self-reported responses if the regression was found to be significant. The latter was because not all those being screened in future will be able to provide a self-report score, though they will nearly always be capable of giving a behavioural response. Only participants with no missing data within each respective category had total scores calculated, in order to ensure the most reliable dataset possible. These overall scores were assessed via ROC analysis to identify an appropriate cut-point and the corresponding sensitivity and specificity.

Subsequently, hierarchical regression was run that included age, gender, and total score on the behavioural screen to predict the person's self-reported AQ score. For those participants with parental/informant raters, the regression was run again but instead substituting the self-reported AQ score for parental/informant AQ.

Reliability was assessed by internal consistency on the behavioural screen items, both when all are scored in a positive direction and separately if any were indicated as needed to be reverse scored. Interrater reliability was assessed on the scoring of the observational scores between the researcher and their primary supervisor, using Kappa for each item. These results were interpreted according to the guidance by Landis and Koch (1977).

All analyses were run in SPSS Statistics Version 24. Relevant syntax, outputs, and datafiles are available in the ESM (Chapter 6 / Analysis).

6.4 Results

6.4.1 Power analyses for ROC

A power analysis for ROC was run using the assumptions of .80 power (Beta) and a 50:50 ratio between autistic and non-autistic participants. This showed that to achieve a ROC value of .70, 31 participants would be required in each group. Given that ROC values were to be calculated for people with learning disability independent of those without learning disability, this meant that the following number of participants were required for the analyses: 31 autistic participants without learning disability, 31 comparable non-autistic controls, 31 autistic participants with co-occurring learning disability, and 31 comparable non-autistic controls with learning disability. Comparing this analysis to the participants included here, only the autistic group was sufficiently powered, while the autistic with learning disability ($N = 22$), potentially autistic ($N = 16$), and all control participant (no learning disability $N = 22$; with learning disability $N = 19$) groups were slightly underpowered.

6.4.2 Attention task scores

The attention task appeared to be an ineffective test of who was and was not attending to the stimuli, as discussed in 6.3.3.2. As such, only the mean score of the task is reported here, and no further analyses of this data will be carried out. The mean score was 23.19 ($SD = 4.09$), the lowest reported score was 0 and the highest was 27 (of a maximum of 27 items).

6.4.3 Binomial Regression

Three regressions were run, to predict diagnostic status (i.e. whether the person is reported as being autistic or not). In these regressions, the outcome variable was always the person's diagnosis (either autistic or non-autistic) and the predictor variables were always of one type, either the self-reported responses, observed responses, or combined responses to each of the behavioural screen stimuli. For each, 27 variables were entered as predictor variables, with scores for the responses to the Neutral and Quiet stimuli being omitted.

Before commencing this series of regressions, collinearity was checked. No correlation, or VIF value caused concern. Average VIF values are shown in Table 50 and the full VIF values and correlations are available in the ESM (Chapter 6 / Analysis / Output).

Table 50: Mean Variance Inflation Factor (VIF) from each group of predictor variables, present in each regression.

	Mean *(SD)* VIF
Self-Reported Only	1.93 (0.44)
Observed Only	1.68 (0.27)
Combined	1.87 (0.39)

Self-reported Responses Only

In the binomial logistic regression, to predict diagnosis, the difference between the null model and that which included the 27 predictor variables was found to be non-significant, see Table 51 (note that the 'Point' stimuli are not included here, due to no variation within the scores). Table 52 shows that the model was able to correctly identify 95.6% of the autistic participants correctly while identifying a lower proportion, 32.4%, of the non-autistic participants correctly. Only Disgust (M) and Adrenaline (F) were found to be significant contributors to the model, see Table 53.

Table 51: Model fit statistics for the self-reported responses binomial logistic regression.

	Chi Square (df)	*p*	-2 Log likelihood	Nagelkerke R^2
Model 1	21.644(25)	.656	124.027	.232

Table 52: Classification table for the self-reported responses binomial logistic regression.

		Predicted		
		Autistic	Non-autistic	Percentage Correct
Actual	Autistic	86	4	95.6
	Non-autistic	23	11	32.4
N = 125	Overall			78.2

Table 53: Variables in the self-reported responses equation and their respective values.

Stimuli*	B	Wald(df)	*p*	Exp(B)
Fear (M)	.995	1.495(1)	.221	2.704
Disgust (F)	.064	.005(1)	.943	1.066
Laugh (F)	-.868	2.030(1)	.154	.420
Sad (M)	.212	.101(1)	.751	1.236
Crowd	-.997	2.114(1)	.146	.369
Sad (F)	.902	1.477(1)	.224	2.464
Push (F)	-.513	.573(1)	.449	.599
Angry (M)	-.508	.440(1)	.507	.602
Disgust (M)	3.028	6.784(1)	.009	20.666
Itch (M)	-.355	.260(1)	.610	.701
Fear (F)	-.059	.008(1)	.928	.943
Angry (F)	-.663	.629(1)	.428	.515
Distract (F)	.190	.073(1)	.787	1.209
Laugh (M)	1.340	3.043(1)	.081	3.136
Push (M)	.381	.398(1)	.528	1.463
Pain (M)	.438	.258(1)	.612	1.549
Yawn (F)	-.237	.089(1)	.766	.789
Adrenaline (M)	1.340	3.467(1)	.063	3.820
Pain (F)	-.232	.034(1)	.855	.793
Happy (M)	-.614	.909(1)	.340	.541
Adrenaline (F)	-1.608	4.670(1)	.031*	.200
Itch (F)	-.160	.044(1)	.834	.852
Distract (M)	.272	.126(1)	.722	1.312
Happy (F)	.911	2.255(1)	.133	2.487
Yawn (M)	-.077	.008(1)	.929	.926
Constant	-1.305	8.186(1)	.004	.271

*Note. M = Male, F = Female

Observed Only Responses

In the binomial logistic regression to predict diagnoses, using only scores for observed responses, the difference between the null model and that which included the 27 predictor variables was again found to be non-significant (see Table 54).

While this model was able to correctly identify 95.7% of the autistic participants, it was unable to identify 64.9% of the non-autistic participants (see Table 55). No items were found to be significant contributors to this model (see Table 56).

As the model itself was non-significant, nor were any individual items significant predictors, a total score was created that reverse scored all items with a negative beta (Table 56) and summed the values.

Table 54: Model fit statistics for the observed responses binomial logistic regression.

	Chi Square (df)	*p*	-2 Log likelihood	Nagelkerke R^2
Model 1	27.07(27)	.460	125.687	.273

Table 55: Classification table for the observed responses binomial logistic regression.

		Predicted		
		Autistic	Non-autistic	Percentage Correct
Actual	Autistic	89	4	95.7
	Non-autistic	23	13	36.1
N = 129	Overall			79.1

Table 56: Variables in the observed responses equation and their respective values.

Stimuli*	B	Wald(df)	*p*	Exp(B)
Fear (M)	-1.602	1.058(1)	.304	.201
Disgust (F)	.104	.008(1)	.930	1.109
Laugh (F)	-.403	.393(1)	.531	.668
Sad (M)	1.339	1.155(1)	.282	3.814
Crowd	-.011	.000(1)	.991	.989
Sad (F)	-1.776	1.496(1)	.221	.169
Push (F)	1.636	3.530(1)	.060	5.132
Angry (M)	-.394	.042(1)	.838	.674
Point (M)	-.189	.047(1)	.829	.828
Disgust (M)	3.054	2.511(1)	.113	21.197
Itch (M)	-.046	006(1)	.937	.955
Fear (F)	-1.011	.702(1)	.402	.364
Angry (F)	-20.791	.000(1)	.999	.000
Distract (F)	.655	.476(1)	.490	1.925
Laugh (M)	-.212	.106(1)	.745	.809
Push (M)	.018	.000(1)	.986	1.019
Pain (M)	1.879	1.175(1)	.278	6.544
Yawn (F)	.033	.001(1)	.970	1.034
Adrenaline (M)	-.258	.180(1)	.671	.773
Pain (F)	.216	038(1)	.845	1.241
Point (F)	-.169	.038(1)	.845	.845
Happy (M)	1.023	2.844(1)	.092	2.780
Adrenaline (F)	.312	.250(1)	617	1.366
Itch (F)	-.244	.178(1)	673	.784
Distract (M)	-1.130	.824(1)	.364	.323
Happy (F)	.193	.084(1)	772	1.213
Yawn (M)	-1.391	1.217(1)	.270	.249
Constant	-.941	5.297(1)	.021	.390

*Note. M = Male, F = Female

Combined Responses

In the binomial logistic regression using the Combined Responses, that takes into account both self-reported and observed reactions, the analysis found that the difference between the null model and that which contained the predictor variables was non-significant (see Table 57).

The model was able to correctly identify 94.3% of the autistic participants, but only 45.2% of the non-autistic group (see Table 57). Although the model overall was found to be non-significant, the Disgust (M) item was found to be a significant contributor to the model, but was only associated with a very small increased likelihood of a response being in a particular direction, leading to the person being autistic (see Table 58).

Accordingly, 2 total scores were created here. Firstly, a total combined score that included all items entered into this regression here and secondly, a short version which only included the 1 significant contributor. Items were reverse scored according to their beta value in Table 59.

Table 57: Model fit statistics for the combined responses binomial logistic regression.

	Chi Square (df)	*p*	-2 Log likelihood	Nagelkerke R^2
Model 1	31.455(27)	.253	104.451	.342

Table 58: Classification table for the combined responses binomial logistic regression.

		Predicted		
		Autistic	Non-autistic	Percentage Correct
Actual	Autistic	82	5	94.3
	Non-autistic	17	14	45.2
	Overall			81.4

Table 59: Variables in the combined responses equation and their respective values.

Stimuli*	B	Wald(df)	*p*	Exp(B)
Fear (M)	.126	.020(1)	.887	1.134
Disgust (F)	-1.580	1.963(1)	.161	.206
Laugh (F)	-1.165	2.427(1)	.119	.312
Sad (M)	.820	1.177(1)	.278	2.270
Crowd	.084	.013(1)	.908	1.088
Sad (F)	-.408	.205(1)	.650	.665
Push (F)	-.220	.097(1)	.756	.803
Angry (M)	-.148	.028(1)	.886	.862
Point (M)	-.637	.499(1)	.480	.529
Disgust (M)	3.370	8.773(1)	.003*	29.077
Itch (M)	-.308	.203(1)	.652	.735
Fear (F)	.285	.150(1)	.699	1.330
Angry (F)	-.401	.187(1)	.666	.670
Distract (F)	-.707	.768(1)	.381	.493
Laugh (M)	.220	.082(1)	.775	1.247
Push (M)	1.167	3.036(1)	.081	3.213
Pain (M)	.650	.710(1)	.399	1.915
Yawn (F)	.280	.105(1)	.746	1.323
Adrenaline (M)	-1.263	2.978(1)	.084	.283
Pain (F)	-.101	.008(1)	.930	.904
Point (F)	.382	.144(1)	.704	1.465
Happy (M)	.296	.145(1)	.703	1.344
Adrenaline (F)	.431	.385(1)	.535	1.539
Itch (F)	-.873	1.579(1)	.209	.418
Distract (M)	.472	.334(1)	.563	1.603
Happy (F)	.916	1.251(1)	.209	2.499
Yawn (M)	-.797	.709(1)	.400	.451
Constant	.525	.703(1)	.402	.592

*Note. M = Male, F = Female

6.4.4 **ROC Analysis**

A ROC analysis was conducted using the total scores created, based upon the results from the regression analyses for observed, combined, and combined short. For all ROC analyses, scores equal to or lower than the chosen cut-point indicated a higher chance of the person being autistic. Different numbers of participants were found in each ROC analysis due to only complete datasets for each response type being used. The means and standard deviations of these total scores for their respective groups are reported in Table 60.

For every ROC analysis performed, on either the whole sample or the learning disability only group, sensitivity was found to be higher than specificity. For the whole sample, the sensitivity values for the observed and combined responses met the 70-80% sensitivity criteria outlined by Glascoe (2005) as being required for a good screening instrument, yet neither met the 80% criteria for specificity. For the learning disability only sample, the observed responses met this criterion for sensitivity, but the combined responses fell slightly short. Neither met the criteria for specificity. The 1-item short score showed high sensitivity but scored very low on specificity, this was true for the whole sample and the learning disability only subsample. Details about AUC, sensitivity, and specificity at the optimal cut-point can be found in Table 60 and full details are available in the ESM (Chapter 6 / Analysis / Output).

Table 60: Results of ROC analyses by sample and response type.

	Whole sample*			Learning Disability Only		
	Observed Responses (*N* = 129)	Combined Responses (*N* = 118)	Combined Short (*N* = 134)	Observed Responses (*N* = 35)	Combined Responses (*N* = 30)	Combined Short (*N* = 38)
Mean *(SD)*	13.66 (1.79)	13.43 (2.00)	.12 (.33)	14.03 (1.76)	14.27 (2.08)	.11 (.31)
Area Under Curve (AUC)	.69	.69	.58	.72	.71	.56
Optimal cut-point (≥)	14.5	14.5	0.5	14.5	14.5	0.5
Sensitivity	.74	.81	.93	.72	.68	.95
Specificity	.58	.45	.24	.71	.57	.17

*Note. (exc. 'potentially autistic participants)

6.4.5 **AQ Hierarchical Regressions**

Participants who had provided AQ scores, including those who reported that they were possibly autistic but not diagnosed, were selected for the next stage of analysis. First, correlations were run to assess the relationship between the total observed score, the total combined score, and each participant's self-reported AQ score. As aforementioned, participants are always able to provide observable responses but are not always able to provide self-reported responses; accordingly, the effectiveness of self-reported scores alone was not assessed here. These correlations were found to be non-significant (see Table 61) and due to these non-significant results, the hierarchical regressions were not run.

Table 61: Self-reported short-form AQ scores correlated with observed and combined behavioural screen scores.

	Observed Score $M = 13.83$ ($SD = 1.80$) ($N = 96$)	Combined Score $M = 13.66$ ($SD = 2.13$) ($N = 86$)
AQ-10 Score $M = 5.95$ ($SD = 2.47$)	$r = -.16$, $p = .127$	$r = -.16$, $p = .134$

Participants with informant full-scale AQ scores were selected, again including those who were possibly autistic. Correlations were run between the observed score, combined score, and the informant AQ score. Both correlations were found to be significant (see Table 62).

Table 62: Informant AQ scores correlated with observed and combined behavioural screen scores.

	Observed Score $M = 13.37$ ($SD = 1.81$) ($N = 93$)	Combined Score $M = 13.16$ ($SD = 1.70$) ($N = 88$)
Informant AQ Score $M = 34.50$ ($SD = 8.29$)	$r = -.27$, $p = .010$	$r = -.28$, $p = .008$

The behavioural screen scores were entered into a hierarchical regression. Block 1 contained age and gender and block 2 either the observed or combined score. Gender was entered due to its known

influence on AQ scores, whereby males tend to score higher than females on the scale (Ruzich, Allison, Smith, et al., 2015). While the influence of age on AQ is less marked (Ruzich, Allison, Chakrabarti, et al., 2015), the sample here contained a wide age range and it was entered in order to control for any potential effect.

The data was checked before running the analysis. Scatter and QQ plots of the residuals were inspected for homoscedasticity and distribution, Cook's distance for outliers, and correlations and VIF for multicollinearity. No assumptions of the regression were violated.

The model including the observed score showed significant change when compared to the block containing only age and gender, F change (1, 89) = 7.624, p = .007, R^2 change = .071. This second model was found to be significant, F(3, 89) = 5.896, p = .001. As can be seen in Table 63, age and observed score contributed significantly; gender did not.

Results were similar for the combined score. When compared to the block containing age and gender, the block including combined score was found to significantly improve the model fit, F change (1, 84) = .067, p = .011, R^2 change = .067. This second model was shown to be significant, F(3, 84) = 5.635, p = .001. Age and combined score both contributed significantly to the model; gender did not (see Table 63).

Table 63: Final regression model statistics for the observed and combined models.

Model and R Values	Item	B	*t*	*p*	VIF
Observed Model R^2 = .166	Constant	54.210	9.112	< .001	
	Age	-.185	-3.140	.002	1.023
	Gender	-.195	-.124	.902	1.041
	Observed Score	-1.175	-2.761	.007	1.020
Combined Model R^2 = .168	Constant	55.352	8.354	< .001	
	Age	-.181	-2.840	.006	1.024
	Gender	-.879	-.532	.596	1.018
	Combined Score	-1.215	-2.599	.011	1.010

6.4.6 **Reliability**

Cronbach's alpha showed that when some of the items were reverse scored, in line with the findings of the regression analyses, a very low internal consistency was found. However, when all items were scored with a response denoted by 1 and no response denoted by 0, internal consistency was far higher, see Table 64. This indicated it may be possible that the forward and reverse scored items had separate groupings. To check this, Cronbach's alpha was run on the forward and reverse scored items separately for the observed and combined responses. These results showed that when they are assessed separately, internal consistency is notably higher. The combined responses also appear to show higher internal consistency than the observed responses. See Table 65 for details.

Table 64: Internal consistency for observed and combined responses, with and without reverse scoring.

	Observed responses (N = 144)	Combined responses (N = 132)	Observed responses (No rev. score) (N = 144)	Combined responses (No rev. score) (N = 132)
Cronbach's alpha	-.048	-.342	.767	.853

Table 65: Internal consistency for observed and combined responses, calculated for forward and reverse scored items separately.

	Observed Responses		Combined Responses	
	Forward scored (N = 144)	Reverse scored (N = 145)	Forward scored (N = 138)	Reverse scored (N = 132)
Cronbach's alpha	.668	.601	.747	.713

The Kappa analyses indicated that the 2 raters generally showed moderate or better agreement on many of the items (Table 66), however they showed no agreement on 4 items: Angry (M), Disgust (M), Pain (F) and Pain (M). Adrenaline, Happy, Laughter and Yawn stimuli were the types that were generally agreed upon between the 2 raters (see Table 67).

Table 66: Interrater agreement stratified by level.

Level of agreement (Landis & Koch, 1977)	Number of items within range
No agreement (< 0)	4
Slight (0 — .20)	-
Fair (.21 — .40)	2
Moderate (.41 — .60)	7
Substantial (.61 — .80)	7
Perfect (.81 — 1.0)	7

Table 67: Interrater reliability of each item.

Item	Kappa	*p*
Adrenaline (F; Item 1)	.774	< .001
Adrenaline (M; Item 14)	.788	< .001
Angry (F; Item 2)	.410	.007
Angry (M; Item 2)	.239	.051
Disgust (F; Item 3)	.650	< .001
Disgust (M; Item 16)	-.061	.678
Distract (F; Item 4)	.868	< .001
Distract (M; Item 17)	.680	< .001
Fear (F; Item 5)	.611	< .001
Fear (M; Item 18)	.503	.002
Happy (F; Item 6)	.855	< .001
Happy (M; Item 19)	.928	< .000
Itch (F; Item 7)	.855	< .001
Itch (M; Item 20)	.714	< .001
Laugh (F; Item 8)	1.00	< .001
Laugh (M; Item 21)	.747	< .001
Pain (F; Item 9)	.094	.240
Pain (M; Item 22)	.188	.088
Point (F; Item 10)	.440	.005
Point (M; Item 23)	.355	.014
Push (F; Item 11)	.448	.004
Push (M; Item 24)	.410	.007
Sad (F; Item 12)	.440	.005
Sad (M; Item 25)	.243	.049
Yawn (F; Item 13)	1.00	< .001
Yawn (M; Item 26)	.887	< .001
Crowd (Item 27)	.529	.001

6.5 Discussion

6.5.1 Results summary

The analysis presented here showed mixed results and many of the null results were likely attributable to low sample size, which will be revisited in the limitations section in more detail. The regression analyses indicated that the behavioural screening tool, in its current form, was unable to discriminate between autistic and non-autistic participants. This implied that the autistic and non-autistic control participants in the current study reported emotional contagion and were observed mimicking behaviours at comparable levels. As none of the regression models were found to be significant, it was not possible to reduce the number of items included in the behavioural screen to only those which discriminated between the groups. The Beta values provided by these non-significant regressions were, however, used to indicate the direction of influence of each item in the behavioural screening tool. Some items were reverse scored and total scores for both the observed and combined behavioural screening tool outcomes were created. A self-report only total score was not created here as not everyone who might use such a tool in the future would be able to provide a self-report of their responses. These scores were entered into a series of ROC curve analyses. Acceptable sensitivity but lower specificity were found for both the whole sample (excluding anyone in the potentially autistic group) and the learning disability specific sample, with the latter group showing slightly better AUC. The combined score for the single significant item (Disgust, Male) was entered into the ROC analysis. The results were notably poorer than its full-length counterpart.

The behavioural screen total scores were used in a series of hierarchical regressions to predict both participants' self-rated short-form, and informant-rated long-form, AQ scores. The addition of the behavioural screening tool scores to the model predicting self-rated AQ scores was non-significant. Though, their addition was a significant improvement to the model predicting informant-rated AQ scores. This indicated that the behavioural screening tool score was in some ways comparable to an informant-rated AQ score. In terms of reliability, the internal consistency when some items were reverse scored was poor, but if no reverse scoring was conducted then the internal consistency was higher. When the internal consistency for forward and reverse scored items were considered separately, these were

notably higher than when the two were combined. As for interrater reliability, most items showed fair, moderate, substantial, and perfect agreement between raters; though a minority showed no agreement.

6.5.2 **Implications for the literature**

The study showed a null difference between autistic and non-autistic participants when using the behavioural screening tool, and no relationship was found between the tool's score and self-rated AQ scores. At the same time, the screening tool scores significantly related to informant AQ scores, which showed that the tool has some potential. This point will be discussed further. These null results were at odds with the majority of current literature that indicate differences in mimicry of emotions between those with and with autism. Past research shows that children with autism are about half as likely to mimic than children without autism (Scambler et al., 2007), that autistic adults are less likely to mimic when viewing videos of emotions (Yoshimura et al., 2015), that mimicry is reduced when participants are not instructed to copy (McIntosh et al., 2006), and that the mimicry reduction is not explained by other factors and is indeed related to autism (Vanvuchelen et al., 2011). Moreover, with regard to specific types of behavioural mimicry, differences between autistic and non-autistic individuals appears to be a robust finding. A reduction of contagiously laughing in response to laughter is frequently reported in autistic people (Helt & Fein, 2016; Reddy et al., 2002), as is a reduction of contagiously yawning when exposed to yawn stimuli (Giganti & Esposito Ziello, 2009; Senju et al., 2007). These previous findings were not mirrored in the results of the current study. A less well-researched area of mimicry is contagious itching, and what little work has been done on this topic implies that autistic people would be more prone to itching that non-autistic participants (Scheub et al., 2020; Schineller, 2018). In the current study, no difference between the groups was found.

However, there are potential reasons why these results may have occurred that have been suggested in previous research. One possibility is the location of the study. A meta-analysis by Edwards (2014) indicates that while mimicry is consistently produced in both laboratory settings and settings familiar to the child (e.g. home, school), the difference is notably reduced in the latter settings. Setting could have influenced the results of the present study. Participants were given a choice of where they would like to

complete the study. Some chose the university, but many opted to do it in a familiar setting, such as at home, school, or an organisation they were involved with.

A further finding, that may be relevant here, is that children with autism are more likely to laugh at cartoons in the presence of their parents (Helt & Fein, 2016), which implies that when autistic people are more comfortable they are more emotionally expressive. In the current study, while some participants undertook the task on their own, others chose to have someone with them during the task. For a number of participants, this was with parents or partners, or for those being tested at schools or external organisations, a familiar staff member. The presence of these people may have made the participants feel more comfortable and potentially more expressive, thereby inadvertently reducing differences in mimicry responses between those with and without autism.

The fixation cross used as part of the behavioural screen task may also have influenced the results. In the study, a fixation cross was displayed on-screen to cue the participants to press the space bar and watch the video when they were ready. As this cross was at the centre of the screen, which was often also the centre of the actor's face in the stimuli, this could have cued some participants to view the actor's face or eyes more than they would have done naturally. Senju et al. (2009) found that when participants were cued to attend the eyes, children with and without autism contagiously yawned to the same extent. Senju and colleagues conclude that eye contact is important for contagion and that by cueing participants to attend the eyes, this leads to no difference between the two groups. Similarly, Helt, Fein, and Vargas (2019) found that when participants are verbally cued to attend the eyes, any differences in contagious yawning or laughter are negated. Accordingly, although the cross was not specifically set to cue the eyes, nor were participants instructed to attend this fixation cross, it is possible that it may have cued participants toward the actor's eyes and face; thus, leading to increased mimicry and contagion in the autistic group.

A further influence may have been the videos which comprised the stimuli. Firstly, stimulus familiarity may have played a role. While sixteen actors were involved in the initial development of the stimuli, not all were included in the final stimuli and some appeared in more than one video clip. This may have led to some level of stimulus familiarity, which increased emotional contagion and mimicry in those with

autism. Helt et al. (2019) shows that when autistic participants are familiar with the person in a stimulus, contagion becomes comparable with non-autistic people.

A second influence may have been the quality of acting. Hennig-Thurau and colleagues (2006) found that how convincing an actor was, moderated the effect of emotional contagion, as measured by self-reported contagion. It was shown that less convincing acting led to less emotional mimicry taking place. Similarly, Juslin, Harmat, and Laukka (2018) found that when participants were exposed to either spontaneous or posed voice clips, measures of the sympathetic nervous system showed that they were more aroused by the spontaneous than posed clips. This indicates that people are able to distinguish between what is real and what is posed, which has been mirrored in past research that shows people are able to perceive when biological motion is acted or posed (Runeson & Frykholm, 1983). As such, it is reasonable to assume that if the actors in the stimuli used in the present study were unconvincing, reactivity would be reduced. The stimuli in this study were developed in multiple stages, including repeated recording sessions and a broad range of actors; some of whom had previous acting experience. In some cases, actors were not actually acting, instead they were indeed yawning or showing pain from receiving a mild electric shock. The videos used were of the highest quality that could be achieved given the time and budget constraints of the project. Unfortunately, it is unknown whether this had any influence as no rating of how convincing these videos actually were, was taken.

While there are many potential reasons why the behavioural screening tool was unable to discriminate between autistic and non-autistic participants or predict self-reported AQ results, it did predict informant-rated AQ scores to a significant degree. It should be noted that this was for both adults and those aged under eighteen. For adults, the informant-rated AQ is not a validated screening tool and scores on a similar version have only been shown to correspond to the self-rated AQ in previous research (Baron-Cohen et al., 2001). For adolescents and children, the tools are both validated screening tools (Auyeung et al., 2008; Baron-Cohen et al., 2006). The children's version, with a cut-point of seventy-six, and a one to four scoring scheme, has a sensitivity and specificity of 95%. The adolescent version shows that with a cut-point of thirty around 89% of autistic adolescents (90.4% Asperger's/high functioning autism; 88.6% autism) score positively on the scale. These findings showed that scores on the under-eighteen

versions are predictive of whether the person is autistic or not. The finding that the behavioural screen was predictive of informant-rated AQ scores implies that it may still have some potential to effectively work as a screening tool. This opens up the possibility of using the behavioural screening tool in place of informant measures. In some sense, this makes the tool similar to the MUSAD (Bergmann et al., 2015), in that it allowed participants who may not have previously been able to undergo screening to participate directly in the screening process. By supporting people to complete screening tools for themselves, this alleviates issues around not being able to identify appropriate informants due to infrequent, or loss of contact with family members, or a high staff turnover meaning staff members are unable to accurately report on the person's behaviours and personality (Bigby, 2008; Butler et al., 2010; Robertson et al., 2005). The study aimed to develop an accessible tool that could achieve this and help inform a diagnosis. This, however, must be viewed in light of the tool not being effective at discriminating between those with and without a reported autism diagnosis. Potential reasons why these null results occurred have been explored in the context of existing literature; though, there were major issues and limitations found within the current study.

6.5.3 **Issues and limitations**

The foremost issue with this study was sample size. As stated in the method, the recruitment and testing for this study was halted due to the Covid-19 outbreak. These events left the study with a sizable sample of autistic people, but the other groups were notably underpowered. The non-autistic control sample was far smaller and did not have sufficient power to reach the recommended threshold. While it contained participants who had been selected as comparable to the autistic participants (i.e. in terms of age and gender), because recruitment was incomplete, the groups were not comparable overall at the point of analysis. The control group was on average older and included a higher percentage of females. The two groups of people with learning disability also had insufficient numbers to enable the study to achieve statistical power. In addition, the group of people with learning disability, but not autism, was older and included a higher percentage of females compared to the group of those with learning disability and autism. The potential effect of these differences on the results, are nonetheless, likely to be relatively small. In respect of age, mimicry exists from an early age (Chartrand & van Baaren, 2009) and studies

have found no, or a very weak influence of age on mimicry in adults (Bailey, Henry, & Nangle, 2009; Sforza et al., 2010). Similarly, researchers have suggested that the overall effect of gender on mimicry is likely to be weak or non-existent depending upon the circumstance and stimuli (Hatfield et al., 2014).

The reduced number of control participants in the study and the fact that the analyses were underpowered, were the most obvious reasons why the behavioural screening tool did not demonstrate the ability to distinguish between those with and without autism. As the point at which data collection could resume was unknown, the researcher took the decision to complete their thesis based on the available data. The researcher aims to resume data collection, following the submission of the thesis and once data collection is allowed to resume. Participants, who agreed to take part but were prevented due to Covid-19 restrictions, will be contacted to rearrange testing. In addition, recruitment of additional participants will be undertaken. Re-analysing the data with a sufficiently powered sample may reveal a different set of results.

A second issue follows this. Due to the non-significant results of the regression, the total score of the behavioural screening tool was based upon the Beta values taken from the non-significant models. With a properly powered study, these analyses may reveal a different pattern of results, identify items that are significant predictors within the model, and provide a different more robust method of creating a total score. This may prove to be more discriminating between autistic and non-autistic individuals.

A further limitation is that the behavioural screening tool was found to have quite low internal reliability. When the reverse scoring was applied, Cronbach's alpha was shown to be negative for both the observed and combined scores. When the internal consistency for the forward and reverse scored items was calculated separately, the internal consistency was found to be much stronger. As with the other analyses, these statistics may differ when further data is added. Additionally, if the regressions are found to be significant and they indicate that some items should be removed, it would be worth performing factor analysis to identify better item groupings which may improve the internal consistency of each factor.

A potentially more important consideration is the interrater reliability in this study, with most, but not all items having acceptable levels of agreement between raters. A marking scheme was developed and used that indicated whether or not a person was reacting in a manner that showed empathy, emotional

contagion, or mimicry. In many cases, applying this was relatively straightforward. The difficulty occurred when responses were less overt or when the participant responded in animated ways, even during the Neutral stimuli. The items on which there was no agreement were both of the Pain stimuli, and the male Angry and Disgust stimuli. These are all more difficult to score than some of the other items, such as Yawning or Laughter which have obvious, congruent responses. There are a number of solutions to this issue. The first would be simply removing these items, which may make sense in a context where they are not able to discriminate and do not significantly load into the regression. A second solution is to further clarify the marking criteria and provide training to raters, using specific examples of responses to these stimuli, while introducing the option of discussion between raters to resolve disagreements.

At the same time, multiple raters may be more difficult to introduce if the behavioural screening tool was being used in practice by either clinicians or teachers, as it would make it more time and resource intensive to use. One possible solution is to use a computer algorithm to code responses. At present, programmes exist that can identify emotions from people's facial expressions (Microsoft, 2020) and body language (Zhang, Zhang, Neoh, Mistry, & Hossain, 2015). There are also commercial products that can be purchased for this purpose, although the relative cost of these tools per user would be quite high. These programmes are not currently designed to identify and code behavioural responses, such as itching or yawning, and therefore, a custom solution would be required. Using commercially available products and customised solutions would have significant cost implications but would likely make the behavioural screening tool far more robust and user-friendly.

With a number of participants, attention was a concern. As the attention task appeared ineffective, there was no robust measure of attention in this study. As such, there was no way of telling how closely the participant watched the videos. Some participants appeared visually distracted, as they would look away while the video was playing or begin talking about an unrelated subject. Others would look at the screen and appear unmoved and disinterested in what was being displayed. When reviewing and coding the videos, without the attention task, there was no clear way of assessing whether the person had properly attended the video.

A final issue to note is the quality of the videos of participants which were obtained. The study relied upon the webcam built into the testing laptop to capture participants' responses. The decision was made to use this as it could encode and label separate files for each emotion for each participant, which made analysis possible. For the majority of the videos, this was perfectly acceptable, and the participant's reaction was very clear. Unfortunately, for some participants this was not the case. Due to the study taking place at multiple testing sites and at different times of day, the lighting in some videos made some participants' responses more difficult to see. Moreover, these poor lighting conditions affected the camera quality by making the recorded images blurry and in extreme cases, affecting the camera's framerate, making videos appear too fast or 'jumpy.' As discussed, data from fourteen participants could not be used due to this recording issue. One solution would be to only collect data under conditions that were suitable for recording, although this would be unlikely to reflect the conditions under which the tool would be used in practice. Another solution is to use a higher quality webcam, that was robust enough to withstand regular use in environments such as schools. Any solution will represent a compromise between better quality recordings versus cost and ease of use.

A related issue was that, at times, participants moved out of view of the camera or possibly performed contagious reactions outside of the camera's field of view; for example, scratching their leg in response to the Itch stimuli. As found by Papoiu et al. (2011), itching is contagious but is not spatially specific and the itch response can occur anywhere on the other person's body. To some degree this is unavoidable as the camera was focussed on the person's face and upper torso. If it was further away it would likely miss some of the smaller facial reactions. Again, this could be addressed with a different camera setup whereby one focusses on the face and another on the person's body, though this would make the tool less user-friendly in practice.

A final issue was that sound was not able to be recorded in the programme. Having sound could have possibly made the coding of the videos easier, as it would have provided more cues to work from. Future research could try to address this by amending the programme to enable sound to be recorded, or the researcher could record sound on a separate device.

6.5.4 **Future research**

Future research on the behavioural screening tool should initially focus on addressing the issues outlined above. The first step will be to recruit more participants to ensure the study is sufficiently powered. The increased dataset should be re-analysed to see if the behavioural screening tool is able to differentiate between those with and without autism. With this data, redundant items (if found) should be removed and dimension reduction should be performed on the remaining items. This would ascertain whether possible subscales exist and to get a true understanding of any item groupings, which would likely influence the internal consistency of the measure. If the tool was found to discriminate, it could be trialled with people who are about to undergo assessment for a diagnosis of autism. This would provide information about whether the tool is predictive of diagnostic status. If so, the results may be useful to clinicians performing diagnostic assessments. Future research could explore the possibility and benefits of using a computer-based scoring algorithm as a way of reducing subjectivity in scoring.

6.5.5 **Conclusion**

Overall, this study has taken the next step in the development of a behavioural screening tool, a development that is by no means complete. Using observed responses, and responses that combine both what is observed and what is self-reported, the study found that the behavioural screening tool was unable to accurately distinguish between autistic and non-autistic people, nor was it able to predict self-reported AQ scores. The most likely reason for the null results was that the study was under-powered, due to recruitment and testing being cut short because of Covid-19 restrictions. This will be addressed in future by collecting further data when restrictions are lifted. Suggestions were also made to improve the reliability of the tool and to address limitations, such as the quality of video recordings taken of participant responses. Despite all this, the behavioural screening tool was able to significantly predict informant-rated AQ scores, which indicates the tool itself may have utility and that it shows real promise if it is refined and developed further through future research.

7.0 Chapter 7: Discussion

This thesis set out to explore autism screening for people with learning disability, before attempting to improve accessibility to screening tools for this group and develop a behaviour-based screening tool. This final chapter summarises the body of work, placing it in the context of existing literature, before moving onto a discussion of the limitations of the research presented and offer directions for future study. Finally, an overall conclusion of the thesis will be given.

7.1 Summary of work

Chapter One

The introductory chapter of the thesis emphasised the high co-occurrence of autism and learning disability (e.g. La Malfa et al., 2004; Matson & Shoemaker, 2009), yet the apparent underdiagnosis of autism in this group (Saemundsen et al., 2010). Having an accurate diagnosis of autism means that appropriate support can be provided (e.g. Koudys, Perry, Ho, & Charles, 2018). A potential solution for underdiagnosis of autism is the use of screening tools to help identify those who would benefit from a full diagnostic assessment (Glascoe, 2005). When implemented, screening tools can help ensure people are referred appropriately, and an accurate diagnosis is achieved more quickly (Allison et al., 2012; McKenzie & Murray, 2015). The challenge, that this thesis aimed to address, is that although there are benefits to screening, it is often inaccessible to people with learning disability. The screening tool recommended by NICE (2012) is the AQ (Baron-Cohen et al., 2001), and while this seems beneficial for many, the requirement of having a high level of literacy ability in order to complete it often excludes people with learning disability.

In this context, the thesis set out the overall aim of improving screening for autism in people with learning disability and three related aims to help accomplish this. The next section is broken down by aim, with discussion of what was accomplished and whether these aims were achieved, before considering the extent to which the overall aim of the thesis was met.

7.1.1 **Aim: To make the AQ more accessible to people with learning disability**

Chapter Two

NICE (2012) recommends using the AQ (Baron-Cohen et al., 2001) as a screening tool for autism, alongside the child and adolescent-specific versions (Auyeung et al., 2008; Baron-Cohen et al., 2006) and their short-form counterparts (Allison et al., 2012). As stated previously, these tools are often not accessible to people with learning disability due to the level of literacy required to complete them. By following published recommendations to improve accessibility (e.g. Townsley et al., 2003), this study adapted the items from the AQ, retaining its use as a self-report instrument, into a format which someone with learning disability would be more likely to be able to answer. Two versions were created, one that retained the four point Likert scale (AccAQ-B) and another which only provided two response options (AccAQ-A). Both long and short versions were created. The aim of this study was to explore the reliability and validity of these adapted tools.

Over five hundred participants completed some form of the AQ. Of these people, over four hundred completed a long-form version and eighty-eight completed the short-form. Overall, the long-form AccAQ total and subscale scores were found to be broadly comparable to the original AQ. Inter-item agreement showed moderate or better agreement for most items, and test-retest reliability was shown to be good. While these findings were promising, the limited number of autistic participants in the sample prevented the discriminant ability of these new tools from being explored. The findings for the short-form AQ were not as strong. Concurrent validity was found to be good between the AccAQ and self-reported AQ scores, but validity was poorer when compared with an informant version. Inter-item agreement was mostly moderate or better; although compared to the long-form, a higher proportion of items showed lower levels of agreement.

The data was split into those who were aged under eighteen, and eighteen years or more. ROC analyses revealed the former failed to reach acceptable levels of sensitivity and specificity. The latter group only achieved acceptable sensitivity. When short-form AccAQ scores were combined with those derived from the long-form AccAQ, the ROC results indicated both good sensitivity and specificity for the AccAQ. Neither AccAQ-A nor AccAQ-B appeared to be psychometrically superior or preferred by participants.

As the number of participants with learning disability in this study was low, no specific analyses could be run to determine the usefulness and psychometric properties of the AccAQs with this group. During the targeted recruitment, it became apparent that, while many people with learning disability could complete the screen, a minority could not. This implied that while the adapted AQs were perhaps more accessible and easier to use, they were not accessible to everyone and still required respondents to have some level of literacy and language skills.

Overall, the aim of this study was partially achieved as more accessible versions of the AQ were developed which had acceptable psychometric properties. Further research, with a larger number of autistic participants, is required to assess the ability of the AccAQ to differentiate between those with and without autism. Also, additional participants with learning disability are required to obtain more detailed information about its usefulness with this specific group. At this point, the AccAQ could be implemented when the original AQ is not accessible to people, in the same way as the children and adolescent versions are employed (Auyeung et al., 2008; Baron-Cohen et al., 2006).

7.1.2 **Aim: To understand what other screening tool options may be available to people with learning disability**

Chapter Three

With the adapted version of the AQ showing promise but still having some issues in terms of accessibility, there was a need to explore if alternative screening tools for autism were available that would be suitable for use with people with learning disability. In Chapter Three, a systematic literature review was conducted, which aimed to compile all of the psychometric data available on autism screening tools that had been used with participants with learning disability. The review provided an overview of the validity and reliability of the tools with this group. Generally, it found that there were indeed tools available for the purpose of screening autism in learning disability, yet relatively few had been specifically designed for this purpose. For a number of these tools, comprehensive and detailed information about reliability and validity was lacking, with many only having one or two studies. Most tools also relied on informants, meaning that the people being screened often had little or no input into the screening process. An exception to this was the MUSAD (Bergmann et al., 2015) which is an

observation-based tool that uses music. This tool, while only having one paper published with people with learning disability, showed considerable promise as a screening instrument. It did, however, require music-therapy related equipment and expertise, which may prove to be a barrier for some services.

The aim of this study was achieved as the review identified and evaluated the psychometric properties of screening tools for autism that are currently available for people with learning disability. Unfortunately, these options were not without limitations: many tools lacked information about, or had poor psychometric properties in relation to people with learning disability; they required the input of third-party informants, which can be a barrier for some with learning disability (Bigby, 2008); or specialist input, which may be a barrier for some services. The review confirmed the need for a behaviour-based screening tool for autism, the development of which was explored in detail in Chapters Five and Six.

7.1.3 Aim: To obtain stakeholder views which would inform the development of a new behavioural screening tool

Chapter Four

Before fully developing the ideas for the behavioural screen, a series of interviews were conducted with: autistic people, some with and some without co-occurring learning disability; parents of autistic people; education professionals and clinicians with experience in autism and/or learning disability. These interviews were conducted to gain a better understanding of existing screening and diagnostic pathways, from multiple points of view. In addition, these were also used to inform the development of the proposed new tool to maximise the likelihood that, if it was effective, it would be used in practice.

The interviews found that many participants had limited knowledge about, and confidence in, identifying autistic people. At the same time, the use of screening tools to support identification was extremely rare and instead, most participants described using gut instinct to spot potentially autistic people. A contributing factor was that those screening tools that they were aware of were felt not to be effective enough to consider using.

Timely diagnosis of autism was seen as important, however, in line with previous research (Crane et al., 2016; Midence & O'Neill, 1999), it was generally felt that the current waiting time for diagnosis and the

diagnostic process itself were too long. Diagnosis of autism was seen as bringing many benefits for the person receiving the diagnosis. These benefits included increased understanding of autistic people and their associated behaviour, both by the person themselves and by others, and being able to access additional support. Research shows that a diagnosis can help people receive appropriate support (Mockett et al., 2011; Ruiz Calzada et al., 2012) and that a wealth of resources exist to help people better understand themselves once they receive a diagnosis (e.g. Faherty, 2014; Farmer, 2020). On the other hand, many participants, both professionals and autistic people, felt that post-diagnostic support was inadequate. Again, this is not the first study to identify frustrations with post-diagnostic support (e.g. Rogers et al., 2016).

Screening was seen as playing a role in facilitating more timely diagnosis, but participants highlighted some of the contextual factors to be considered, if screening were to be used in practice. These included the need for screening to be viewed as a legitimate role of professionals, such as teachers, and the barriers of organisational and individual resistance to change. Participants felt that these could be addressed if any new tool had a clearly identified purpose within the diagnostic pathway, was low cost, time-effective, and practical to use. It was also important to participants that it provided useful information that gave weight to the views of those using it, about whether a person was likely to be autistic. Moreover, participants emphasised the need for any proposed tool to be engaging for those being screened. Research confirms that many of these qualities are important in influencing whether tools are used or not (Richardson et al., 2017). Likewise, social validity (acceptability and satisfaction) has been found to have a major influence on the uptake of evidence-based practice in general (McNeill, 2019).

Chapter Four achieved the aim of obtaining views from a range of stakeholders which would help inform the development of a new behavioural screening tool. As well as identifying wider contextual factors relevant to the development of the tool, participants also provided specific ideas about items that could be included in the proposed screening tool, as outlined in Chapter Six.

7.1.4 **Aim: Proof of concept for development of a new behavioural screening tool**

Chapter Five

The next stage identified the key research and evidence base that would help identify items likely to differentiate between those with and without autism, in order to prove the concept behind a behavioural screening tool. Chapter Five outlines relevant research in relation to the concepts of empathy (Cuff et al., 2016), mimicry (Hatfield et al. 2014), and emotional contagion (Hatfield et al., 1993), before exploring how these constructs may be interlinked (Hess & Fischer, 2013; Neal & Chartrand, 2011; Niedenthal, 2007). Broadly speaking, these concepts are often expressed in different ways or at a reduced level in autistic people, compared to those without autism. These differences appear most notable in relation to cognitive empathy, with emotional contagion showing mixed results (see: Hadjikhani et al., 2009, 2014; Scambler et al., 2007; Song et al., 2019). Further, some specific kinds of behavioural mimicry, including contagiously laughing and yawning, appear to be observed less frequently in autistic people (e.g. Reddy et al., 2002; Senju et al., 2007).

Chapter Five reported on a series of pilot studies that drew on relevant literature, to identify stimuli that may be suitable in a behavioural screening tool. These studies used pre-existing video clips as stimuli and investigated the relationship between participants' self-reported reactions to these and their AQ score.

The aim of Chapter Five, to demonstrate proof of concept for the development of a behavioural screening tool, was achieved. The studies demonstrated that consistent reactions to stimuli occurred, and a machine learning analysis identified stimuli that were the most effective and could predict participants' AQ scores. This confirmed that reactions to certain stimuli could be indicative of autistic traits. A further finding of the pilot studies was that it is possible to create a relatively simple process for marking self-reported responses, which required only limited training and could be applied in a reliable manner. Moreover, the responses that participants reported having and what was observed by an independent researcher were often incongruent. This suggested the need to explore both self-reported and an observed response together, for each item in the proposed behavioural screen.

7.1.5 **Aim: Developing and evaluating the behavioural screening tool**

Chapter Six

Chapter Six, the final empirical study of the thesis, drew on the research and ideas identified in the previous chapters in order to develop and evaluate a behavioural screening tool for autism. Items were refined with some being removed (e.g. videos featuring animals), based on the results of Chapter Five, while others were included, based on research, such as 'itching' (Schineller, 2018). In addition, suggestions from stakeholders who were interviewed in Chapter Four were included: videos showing high adrenaline activities; crowded streets; people being pushed; non-verbal instructions. All items were designed to elicit mimicry, emotional contagion, and empathy. Participants self-reported their responses to the new behavioural screening tool, as well as having their responses recorded on video. Both types of response were scored according to whether they indicated mimicry, emotional contagion, or empathy. The decision to use self-report and video recordings together was to ensure that both the internal state of the person and their outward appearance were taken into account.

The results of this study were mostly non-significant. The behavioural screen was unable to discriminate between autistic and non-autistic people, with and without learning disability. In addition, total behavioural screen scores were unable to predict a person's self-reported AQ score, although it could predict informant-rated AQ score. It was suggested that the most likely reason for these null results was that the study was underpowered. Recruitment and data collection had to be stopped prematurely because of restrictions imposed because of the Covid-19 pandemic.

The results of Chapter Six, based on the data that was available, suggested that the screening tool may have some value, as indicated by its ability to predict informant-rated AQ scores. Further research with an adequately powered study is needed to determine whether the utility of the behavioural screening tool extends beyond this. Viewed in light of the limitations, the findings presented here give cause for optimism that the tool could indeed be effective in the near future.

Taken together, the results of Chapters Five and Six are generally consistent with those found in previous research. Chapter Five found that responses to items, which were designed to elicit responses of mimicry

and contagion, were able to predict AQ score. This shows that responses to these types of stimuli are related to a person's autistic traits. Previous research which investigated differences in mimicry and contagion, in terms of AQ score, has not always found significant results (e.g. Chan & Tseng, 2017; Usui et al., 2013). This is despite differences being found between autistic and non-autistic people in relation to mimicry and contagion (see: Hadjikhani et al., 2009; Scambler et al., 2007; Schineller, 2018; Song et al., 2019) and differences existing between these groups in terms of autistic traits (Baron-Cohen et al., 2001). This disparity, between previous research and the studies outlined here, may be due to the behavioural screening tool including a variety of stimuli, rather than focussing on one specific aspect of mimicry and contagion; which many of the studies that found no relationship between AQ score and contagion did. This suggests that combining responses to a number of video types may be better when attempting to estimate a person's level of autistic traits. Chapter Six found a similar result, but only in respect of informant-rated AQ scores. However, further research is required to confirm this and to determine if specific items on the behavioural screen are more discriminating than others. Future research would consider whether the results of the behavioural screen can go beyond predicting AQ scores, to also accurately identify whether someone is autistic or not.

7.1.6 Overall aim: To improve screening for autism, specifically for people with learning disability

Collectively, the full or partial achievement of the aims for each individual study go some way to addressing the overall aim of the thesis (i.e. improving screening for autism, specifically for people with learning disability). Firstly, the adapted AQ increased the likelihood of the AQ being accessible to some, but not all people with learning disability. Unfortunately, it was not possible to report on the psychometric properties of the AccAQ, specifically in relation to people with learning disability, because of the relatively low number in the sample. Next, the review called attention to several tools that could be considered autism screening instruments which have been used with people with learning disability. The associated psychometric properties of these tools were outlined, thereby enabling clinicians and others to make a more informed choice about the screening tool they use with people with learning disability. Finally, while the behavioural screening tool is unable to be considered a reliable and validated autism

screening tool at present, the results demonstrated proof of concept. With some further research, it is likely it could be developed into an effective and useful screening tool, that does not rely on input from informants, is relatively quick and easy to use, and enables people with learning disability to participate directly in the screening process.

7.2 Limitations and future directions

The limitations of the thesis have been discussed in detail, in the context of each of the studies. The main limitation can be summarised as an inadequate sample size, leading to underpowered studies. This was directly caused by the unexpected restrictions imposed as a result of Covid-19, as participants had been recruited and further testing arranged at the point at which lockdown was introduced.

This impacted different studies in different ways. In Chapter Two, the AccAQ tool was designed to be more accessible to people with learning disability, but the number of people with learning disability in the sample was relatively small, meaning its specific effectiveness with this group could not be established. By contrast, Chapter Six had too few participants in the control group, meaning that the ability to properly evaluate the accuracy of the behavioural screening tool, when discriminating between autistic and non-autistic people, was compromised.

Attempts were made to move this study online, however this was not possible. After trialling various video-calling methods (including Skype, Microsoft Teams, FaceTime, WhatsApp), it was found that due to bandwidth issues the video stimuli and the recordings of participants were disrupted and unclear. Adapting the methodology, to enable participants to view the stimuli without the researcher being present, was considered but rejected. This was because lockdown restrictions would have made it impossible to deliver the laptop, which contained the stimuli, to participants. In addition, past literature indicates that the way in which people view videos is affected by the presence of others (Helt & Fein, 2016). This suggests that the data collected under different circumstances would not be comparable. Collecting further data, using the original method, is planned and will be undertaken when Covid-19 restrictions allow.

A larger data set would also allow for a more detailed exploration of the psychometric properties of the behavioural screening tool. An area that is likely to be of particular importance is whether it demonstrates measurement invariance in relation to gender. Currently, autism research, and therefore the tools developed to screen and diagnose it, are arguably biased towards males (Beggiato et al., 2017; Haney, 2016). This goes some way to explain the male to female imbalance in numbers of people diagnosed as autistic (Loomes, Hull, & Mandy, 2017). Research also suggests that autistic females camouflage stereotypically autistic behaviour more than autistic males, meaning that they moderate behaviour to conform to the conventions of non-autistic social behaviour (Hull et al., 2020; Lai et al., 2017; Rynkiewicz et al., 2016). Considering the results of the behavioural screening tool in the context of other relevant measures, such as the Camouflaging Autistic Traits Questionnaire (Hull et al., 2018), may help address such issues. A larger data set will also potentially allow those items, which are poorer at correctly classifying autistic people, to be identified and removed from the behavioural screening tool, thereby making it quicker and easier to use.

In addition to the collection of further data, future research will explore solutions to the issues of reliability when scoring responses to the behavioural screening tool. As previously discussed, using computer algorithms and machine learning approaches to scoring may be possible and may increase the reliability of the coding. Future research will explore the feasibility of incorporating such approaches.

If subsequent research demonstrates that the behavioural screening tool is effective and accurate, the next stage would evaluate its use in practice, for example in education and clinical settings. The information from the interviews with stakeholders, conducted in Chapter Four, will be helpful here. Some of the potential challenges that are likely to be faced, such as resistance to change, have already been highlighted and can be considered when conducting practice-based research.

Another aspect that became clear in the stakeholder interviews was that many participants felt post-diagnostic support is inadequate. While groups and organisations do exist that provide excellent support

and advice to people (e.g. Autism Information Hub[3]; National Autistic Society[4]) plainly, it is not enough. For screening and subsequent diagnosis to be beneficial, it is important that appropriate support is provided as required. As highlighted by Fischer, Morris and Martines (2014), efforts to identify children with developmental disabilities are only warranted when this can lead to intervention and support for that person. This was articulated well by a clinician in Chapter Four, who pointed out: *'there's not many people who are just curious and just want to see [whether they're autistic] and crack on' (C27).*

People seeking a diagnosis, seldom do so as an end in itself. Screening for autism should be considered in the wider context of existing diagnostic services, pathways and support. As the ultimate aim of the screening tool is to improve the lives of autistic people, an important area for future research is to explore the impact that the behavioural screening tool has on those being screened, in particular whether it facilitates access to better support.

7.3 Overall conclusions

The aim of the research outlined in this thesis was to improve autism screening for people with learning disability. Taking account of the limitations, the research has highlighted existing tools for this group of people and their associated psychometric properties. It has developed more accessible versions of the AQ, which have comparable properties with the corresponding existing validated versions. With this, it has also identified factors that key stakeholders consider important in the development and use of a screening tool for autism. Importantly, it has demonstrated proof of concept for a behavioural screening tool and used this to develop a version of such a tool, that was able to significantly predict participants' informant-rated AQ scores. Though further work is needed to develop this screening tool, these studies have generated novel information and made a significant contribution to the development of new and more accessible screening tools for autism, for people with learning disability.

---

[3] https://www.gateshead.gov.uk/article/3896/Autism-Information-Hub
[4] https://www.autism.org.uk/

8.0 References

Alberto, P., & Troutman, A. (2008). *Applied Behavior Analysis for Teachers* (8th ed.). Columbus OH: Pearson.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist for autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(2), 202–212. https://doi.org/10.1016/j.jaac.2011.11.003

American Psychiatric Association. (2013). DSM-5 intellectual disability fact sheet. *American Psychiatric Association*, 2.

American Psychiatric Association [APA]. (1980). *Diagnostic and statistical manual of mental disorders (DSM-3)* (3rd ed.). American Psychiatric Association.

American Psychiatric Association [APA]. (1987). *Diagnostic and statistical manual of mental disorders (DSM-3, Text Revision)* (3rd, rev. ed.). American Psychiatric Association.

American Psychiatric Association [APA]. (1994). *Diagnostic and statistical manual of mental disorders (DSM-4)* (4th ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association [APA]. (2000). *Diagnostic and statistical manual of mental disorders (DSM-4, Text Revision)*. Washington DC: American Psychiatric Association.

American Psychiatric Association [APA]. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington DC: American Psychiatric Association.

Anderson, J. R., & Meno, P. (2003). Psychological influences on yawning in children. *Current Psychology Letters. Behaviour Brain & Cognition*, *2*(11), 1–6. Retrieved from https://journals.openedition.org/cpl/390

Appleby, J., & Robertson, R. (2016). *Public satisfaction with the NHS in 2015: Results and trends from the British Social Attitudes survey*. Retrieved from https://www.kingsfund.org.uk/sites/default/files/BSA-public-satisfaction-NHS-Kings-Fund-

2015.pdf

Arias, V. B., Gómez, L. E., Morán, M. L., Alcedo, M. Á., Monsalve, A., & Fontanil, Y. (2018). Does quality of life differ for children with autism spectrum disorder and intellectual disability compared to peers without autism? *Journal of Autism and Developmental Disorders*, *48*(1), 123–136. https://doi.org/10.1007/s10803-017-3289-8

Aron, A., Coups, E. J., & Aron, E. N. (2013). *Statistics for Psychology* (Sixth). New Jersey: Pearson.

Arthur, A. R. (1999). Using staff consultation to facilitate the emotional development of adults with learning disabilities. *British Journal of Learning Disabilities*, *27*(3), 93–98. https://doi.org/10.1111/j.1468-3156.1999.tb00096.x

Arun, P., & Chavan, B. S. (2018). Development of a screening instrument for autism spectrum disorder: Chandigarh Autism Screening Instrument. *Indian Journal of Medical Research*, *147*, 369–375. https://doi.org/10.4103/ijmr.IJMR_1968_16

Asperger, H. (1991). Autistic psychopathy in childhood in autism and asperger syndrome. In U. Frith (Ed.), *Autism and Asperger syndrome* (1st ed., pp. 37–92). Cambridge: Cambridge University Press.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the autism spectrum quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460. https://doi.org/10.1016/j.paid.2004.04.022

Autism Information Hub. (2020). Autism Information Hub - Gateshead Council. Retrieved July 15, 2020, from https://www.gateshead.gov.uk/article/3896/Autism-Information-Hub

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism spectrum quotient: Children's version (AQ-Child). *Journal of Autism and Developmental Disorders*, *38*(7), 1230–1240. https://doi.org/10.1007/s10803-007-0504-z

Bailey, P. E., Henry, J. D., & Nangle, M. R. (2009). Electromyographic evidence for age-related differences in the mimicry of anger. *Psychology and Aging*, *24*(1), 224–229.

https://doi.org/10.1037/a0014112

Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, *27*(1), 3–23. https://doi.org/10.2307/3315487

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Boston: MIT Press/Bradford Books.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The autism-spectrum quotient (AQ) - adolescent version. *Journal of Autism and Developmental Disorders*, *36*(3), 343–350. https://doi.org/10.1007/s10803-006-0073-6

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644. https://doi.org/10.2307/3094912

Bartholomew, A. J., & Cirulli, E. T. (2014). Individual variation in contagious yawning susceptibility Is highly stable and largely unexplained by empathy or other known factors. *PLoS ONE*, *9*(3), e91773.

https://doi.org/10.1371/journal.pone.0091773

Beggiato, A., Peyre, H., Maruani, A., Scheid, I., Rastam, M., Amsellem, F., … Delorme, R. (2017). Gender differences in autism spectrum disorders: Divergence among specific core symptoms. *Autism Research*, *10*(4), 680–689. https://doi.org/10.1002/aur.1715

Ben-Itzchak, E., & Zachor, D. A. (2007). The effects of intellectual functioning and autism severity on outcome of early behavioral intervention for children with autism. *Research in Developmental Disabilities*, *28*(3), 287–303. https://doi.org/10.1016/J.RIDD.2006.03.002

Bergmann, T., Sappok, T., Diefenbacher, A., Dames, S., Heinrich, M., Ziegler, M., & Dziobek, I. (2015). Music-based autism diagnostics (MUSAD) – A newly developed diagnostic measure for adults with intellectual developmental disabilities suspected of autism. *Research in Developmental Disabilities*, *43–44*, 123–135. https://doi.org/10.1016/J.RIDD.2015.05.011

Berry, W. D., Feldman, S., & Stanley-Feldman, D. (1985). *Multiple regression in practice (No. 50)*. Sage.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Clinical Psychiatry*, *175*(5), 444–451. https://doi.org/10.1192/bjp.175.5.444

Beykikhoshk, A., Arandjelović, O., Phung, D., Venkatesh, S., & Caelli, T. (2015). Using Twitter to learn about the autism community. *Social Network Analysis and Mining*, *5*(1), 1–17. https://doi.org/10.1007/s13278-015-0261-5

Bigby, C. (2008). Known well by no-one: Trends in the informal social networks of middle-aged and older people with intellectual disability five years after moving to the community. *Journal of Intellectual and Developmental Disability*, *33*(2), 148–157. https://doi.org/10.1080/13668250802094141

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

Bittner, C. A. (2018). The importance of role clarity for development of interprofessional teams. *Journal of Continuing Education in Nursing*, *49*(8), 345–347. https://doi.org/10.3928/00220124-20180718-04

Bizzari, V. (2018). Schizophrenia and Common Sense: A Phenomenological Perspective. In I. Hipólito, J. Gonçalves, & J. G. Pereira (Eds.), *Schizophrenia and Common Sense: Explaining the Relation Between Madness and Social Values* (1st ed., Vol. 12). Cham, Switzerland: Springer International Publishing.

Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders*, *43*(12), 2997–3000. https://doi.org/10.1007/s10803-013-1844-5

Brace, N., Kemp, R., & Snelgar, R. (2016). *SPSS For Psychologists (And Everybody Else)* (Sixth). New York: Routledge.

Bradshaw, J. (2001). Complexity of staff communication and reported level of understanding skills in adults with intellectual disability. *Journal of Intellectual Disability Research*, *45*(3), 233–243. https://doi.org/10.1046/j.1365-2788.2001.00318.x

Bradshaw, J., Koegel, L. K., & Koegel, R. L. (2017). Improving functional language and social motivation with a parent-mediated intervention for toddlers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*(8), 2443–2458. https://doi.org/10.1007/s10803-017-3155-8

Bradshaw, J. W. S., & Paul, E. S. (2010). Could empathy for animals have been an adaptation in the evolution of homo sapiens. *Animal Welfare*, *19*(Supp), 107–112.

Braiden, H. J., Bothwell, J., & Duffy, J. (2010). Parents' experience of the diagnostic process for autistic spectrum disorders. *Child Care in Practice*, *16*(4), 377–389. https://doi.org/10.1080/13575279.2010.498415

Braun, J. M., Daniels, J. L., Kalkbrenner, A., Zimmerman, J., & Nicholas, J. S. (2009). The effect of

maternal smoking during pregnancy on intellectual disabilities among 8-year-old children. *Paediatric and Perinatal Epidemiology*, *23*, 482–491. https://doi.org/10.1111/j.1365-3016.2009.01056.x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bravo Oro, A., Navarro-Calvillo, M. E., & Esmer, C. (2014). Autistic behavior checklist (ABC) and its applications. In *Comprehensive Guide to Autism* (pp. 2787–2798). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-4788-7_164

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., & Bird, G. (2016). Can neurotypical individuals read autistic facial expressions? Atypical production of emotional facial expressions in autism spectrum disorders. *Autism Research*, *9*(2), 262–271. https://doi.org/10.1002/aur.1508

Brugha, T., Cooper, S. A., McManus, S., Purdon, S., Smith, J., Scott, F. J., … Tyrer, F. (2012). Estimating the prevalence of Autism Spectrum Conditions in adults: Extending the 2007 adult psychiatric morbidity survey. London: NHS, The Health and Social Care Information Centre.

Bryson, S. E., Bradley, E. A., Thompson, A., & Wainwright, A. (2008). Prevalence of autism among adolescents with intellectual disabilities. *The Canadian Journal of Psychiatry*, *53*(7), 449–459. https://doi.org/10.1177/070674370805300710

Buell, S., Langdon, P. E., Pounds, G., & Bunning, K. (2020). An open randomized controlled trial of the effects of linguistic simplification and mediation on the comprehension of "easy read" text by people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *33*(2), 219–231. https://doi.org/10.1111/jar.12666

Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatrics*, *168*(8), 721–728. https://doi.org/10.1001/jamapediatrics.2014.210

Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Cohen's kappa. *Epidemiology Biostatistics and Public Health*, *14*(2), e12267-1-e12267-10.

https://doi.org/10.2427/12267

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for cronbach's alpha test: A simple guide for researchers. *Malaysian Journal of Medical Sciences*, *25*(6), 85–99. https://doi.org/10.21315/mjms2018.25.6.9

Butcher, J., & Whissell, C. (1984). Laughter as a function of audience size, sex of the audience, and segments of the short film "Duck Soup." *Perceptual and Motor Skills*, *59*(3), 949–950. https://doi.org/10.2466/pms.1984.59.3.949

Butler, S. S., Simpson, N., Brennan, M., & Turner, W. (2010). Why do they leave? Factors associated with job termination among personal assistant workers in home care. *Journal of Gerontological Social Work*, *53*(8), 665–681. https://doi.org/10.1080/01634372.2010.517236

Byström, K. M., & Persson, C. A. L. (2015). The meaning of companion animals for children and adolescents with autism: The parents' perspective. *Anthrozoos*, *28*(2), 263–275. https://doi.org/10.1080/08927936.2015.11435401

Callahan, K., Hughes, H. L., Mehta, S., Toussaint, K. A., Nichols, S. M., Ma, P. S., … Wang, H.-T. (2017). Social validity of evidence-based practices and emerging interventions in autism. *Focus on Autism and Other Developmental Disabilities*, *32*(3), 188–197. https://doi.org/10.1177/1088357616632446

Callus, A., & Cauchi, D. (2020). Ensuring meaningful access to easy-to-read information: A case study. *British Journal of Learning Disabilities*, *48*(2), 124–131. https://doi.org/10.1111/bld.12306

Cassidy, S., & Rodgers, J. (2017). Understanding and prevention of suicide in autism. *The Lancet Psychiatry*, *4*(6), e11. https://doi.org/10.1016/S2215-0366(17)30162-1

Centers for Disease Control and Prevention. (2020). Screening and diagnosis of autism spectrum disorder for healthcare providers. Retrieved March 9, 2020, from https://www.cdc.gov/ncbddd/autism/hcp-screening.html

Chakraborty, D. P. (2010). Prediction accuracy of a sample-size estimation method for ROC studies.

Academic Radiology, *17*(5), 628–638. https://doi.org/10.1016/j.acra.2010.01.007

Chan, M. H. M., & Tseng, C.-H. (2017). Yawning detection sensitivity and yawning contagion. *I-Perception*, *8*(4), 204166951772679. https://doi.org/10.1177/2041669517726797

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In *The New Unconscious* (pp. 334–361).

Chartrand, T. L., & van Baaren, R. (2009). Chapter 5: Human mimicry. *Advances in Experimental Social Psychology*, *41*, 219–274. https://doi.org/10.1016/S0065-2601(08)00405-X

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(12), 1663–1672. https://doi.org/10.1007/s10803-009-0803-7

Chen, L. (2012). Empathy and its implications in high-functioning individuals with autism spectrum disorder (ASD). *Indiana University Undergraduate Journal of Cogntiive Science*, *7*, 17–21.

Chinn, D., & Homeyard, C. (2017). Easy read and accessible information for people with intellectual disabilities: Is it worth it? A meta-narrative literature review. *Health Expectations*, *20*(6), 1189–1200. https://doi.org/10.1111/hex.12520

Chokroborty-Hoque, A., Alberry, B., & Singh, S. M. (2014). Exploring the complexity of intellectual disability in fetal alcohol spectrum disorders. *Frontiers in Pediatrics*, *2*(Aug). https://doi.org/10.3389/fped.2014.00090

Church, C., Alisanski, S., & Amanullah, S. (2000). The social, behavioral, and academic experiences of children with asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, *15*(1), 12–20. https://doi.org/10.1177/108835760001500102

Cohen, B. H. (2008). *Explaining Psychological Statistics* (Third). Hoboken, New Jersey, New Jersey: John Wiley & Sons, Inc.

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives*

of General Psychiatry, *60*(5), 524–530. https://doi.org/10.1001/archpsyc.60.5.524

Cooper, S.-A., McLean, G., Guthrie, B., McConnachie, A., Mercer, S., Sullivan, F., & Morrison, J. (2015). Multiple physical and mental health comorbidity in adults with intellectual disabilities: Population-based cross-sectional analysis. *BMC Family Practice*, *16*(1), 110. https://doi.org/10.1186/s12875-015-0329-3

Cortes, M. J., Orejuela, C., Castellvi, G., Folch, A., Rovira, L., Salvador-Carulla, L., … Martinez-Leal, R. (2018). Psychometric properties of Spanish adaptation of the PDD-MRS scale in adults with intellectual developmental disorders: The EVTEA-DI scale. *Journal of Autism and Developmental Disorders*, *48*(5), 1566–1578. https://doi.org/10.1007/s10803-017-3416-6

Cottrell, L. S., & Dymond, R. F. (1949). The empathic responses: A neglected field for research. *Psychiatry*, *12*(4), 355–359. https://doi.org/10.1080/00332747.1949.11022747

Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2016). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism*, *20*(2), 153–162. https://doi.org/10.1177/1362361315573636

Critical Appraisal Skills Programme (CASP). (2018). CASP Diagnostic Test Study Checklist. Retrieved from https://casp-uk.net/wp-content/uploads/2018/03/CASP-Diagnostic-Checklist-2018_fillable_form.pdf

Croen, L. A., Zerbo, O., Qian, Y., Massolo, M. L., Rich, S., Sidney, S., & Kripke, C. (2015). The health status of adults on the autism spectrum. *Autism*, *19*(7), 814–823. https://doi.org/10.1177/1362361315577517

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Czech, H. (2018). Hans Asperger, national socialism, and “race hygiene” in Nazi-era Vienna. *Molecular Autism*, *9*(1), 29. https://doi.org/10.1186/s13229-018-0208-6

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., … Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. https://doi.org/10.1038/nn1421

Dawson, G. (2016). Why it’s important to continue universal autism screening while research fully examines its impact. *JAMA Pediatrics*, *170*(6), 527–528. https://doi.org/10.1001/jamapediatrics.2016.0163

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Volkmar, F., & Minderaa, R. (2003). Measuring pervasive developmental disorders in children and adolescents with mental retardation: A comparison of two screening instruments used in a study of the total mentally retarded population from a designated area. *Journal of Autism and Developmental Disorders*, *33*(6), 595–605. https://doi.org/10.1023/B:JADD.0000005997.92287.a3

Deputte, B. L. (1994). Ethological study of yawning in primates. I. Quantitative analysis and study of causation in two species of old world monkeys (Cercocebus albigena and Macaca fascicularis). *Ethology*, *98*(3–4), 221–245. https://doi.org/10.1111/j.1439-0310.1994.tb01073.x

Derks, O., Heinrich, M., Brooks, W., Sterkenburg, P., McCarthy, J., Underwood, L., & Sappok, T. (2017). The social communication questionnaire for adults with intellectual disability: SCQ-AID. *Autism Research*, *10*(9), 1481–1490. https://doi.org/10.1002/aur.1795

DiGuiseppi, C., Hepburn, S., Davis, J. M., Fidler, D. J., Hartway, S., Lee, N. R., … Robinson, C. (2010). Screening for autism spectrum disorders in children with down syndrome: Population prevalence and screening test characteristics. *Journal of Developmental and Behavioral Pediatrics*, *31*(3), 181. https://doi.org/10.1097/DBP.0b013e3181d5aa6d

Dillenburger, K., Keenan, M., Doherty, A., Byrne, T., & Gallagher, S. (2010). Living with children diagnosed with autistic spectrum disorder: Parental and professional views. *British Journal of Special Education*, *37*(1), 13–23. https://doi.org/10.1111/j.1467-8578.2010.00455.x

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, *19*(6), 643–647. https://doi.org/10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*(1), 86–89. https://doi.org/10.1111/1467-9280.00221

Donvan, J., & Zucker, C. (2016, January). The Early History of Autism in America. *Smithsonian Magazine*. Retrieved from https://www.smithsonianmag.com/science-nature/early-history-autism-america-180957684/

Duchan, E., & Patel, D. R. (2012). Epidemiology of autism spectrum disorders. *Pediatric Clinics of North America*, *59*(1), 27–43. https://doi.org/10.1016/j.pcl.2011.10.003

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research*, *7*(3), 363–380. https://doi.org/10.1002/aur.1379

Emerson, E., Einfeld, S. L., & Ellis, L. A. (2011). Comorbidity of intellectual disability and mental disorder in children and adolescents: A systematic review. *Article in Journal of Intellectual & Developmental Disability*, *36*(2), 137–143. https://doi.org/10.1080/13668250.2011.572548

Emerson, E., & Hatton, C. (1996). Deinstitutionalization in the UK and Ireland: Outcomes for service users. *Journal of Intellectual and Developmental Disability*, *21*(1), 17–37. https://doi.org/10.1080/13668259600033021

Faherty, C. (2014). *Autism...What Does It Mean To Me?* (2nd ed.). Arlington, USA: Future Horizons Inc.

Fairthorne, J., Langridge, A., Bourke, J., & Leonar, H. (2013). Pre-Existing Differences in Mothers of Children with Autism Spectrum Disorder and/or Intellectual Disability: A Review. In M. Fitzgerald (Ed.), *Recent Advances in Autism Spectrum Disorders - Volume I* (Vol. 1, pp. 387–423). InTech. https://doi.org/10.5772/54488

Fajardo, I., Ávila, V., Ferrer, A., Tavares, G., Gómez, M., & Hernández, A. (2014). Easy-to-read texts for students with intellectual disability: Linguistic factors affecting comprehension. *Journal of Applied Research in Intellectual Disabilities*, *27*(3), 212–225. https://doi.org/10.1111/jar.12065

Farmer, S. (2020). Learning to Love Yourself . Retrieved August 26, 2020, from https://www.aane.org/learning-to-love-yourself/

Faso, D. J., Sasson, N. J., & Pinkham, A. E. (2014). Evaluating posed and evoked facial expressions of emotion from adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(1), 75–89. https://doi.org/10.1007/s10803-014-2194-7

Feneran, A., O'Donnell, R., Press, A., Yosipovitch, G., Cline, M., Dugan, G., … Shively, C. (2013). Monkey see, monkey do: Contagious itch in nonhuman primates. *Acta Dermato Venereologica*, *93*(1), 27–29. https://doi.org/10.2340/00015555-1406

Ferraioli, S. J., & Harris, S. L. (2009). The impact of autism on siblings. *Social Work in Mental Health*, *8*(1), 41–53. https://doi.org/10.1080/15332980902932409

Fischer, V. J., Morris, J., & Martines, J. (2014). Developmental screening tools: Feasibility of use at primary healthcare level in low- and middle-income settings. *Journal of Health, Population and Nutrition*, *32*(2), 314–326. https://doi.org/10.3329/jhpn.v32i2.2625

Fletcher-Watson, S., & Happé, F. (2019). *Autism: A New Introduction to Psychological Theory and Current Debate* (1st ed.). Oxon.

Flick, U. (2018). *An Introduction to Qualitative Research* (6th ed.). London: SAGE Publications.

Forbes, P. A. G., Pan, X., & Antonia, A. F. (2016). Reduced mimicry to virtual reality avatars in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(12), 3788–3797. https://doi.org/10.1007/s10803-016-2930-2

Forrester-Jones, R., Carpenter, J., Coolen-Schrijner, P., Cambridge, P., Tate, A., Beecham, J., … Wooff, D. (2006). The social networks of people with intellectual disability living in the community 12 years after resettlement from long-stay hospitals. *Journal of Applied Research in Intellectual Disabilities*, *19*(4), 285–295. https://doi.org/10.1111/j.1468-3148.2006.00263.x

Franklin, R. G., Nelson, A. J., Baker, M., Beeney, J. E., Vescio, T. K., Lenz-Watson, A., & Adams, R. B. (2013). Neural responses to perceiving suffering in humans and animals. *Social Neuroscience*, *8*(3),

217–227. https://doi.org/10.1080/17470919.2013.763852

Frith, U., & Happé, F. (2005). Autism spectrum disorder. *Current Biology*, *15*(19), 786–790. https://doi.org/10.1016/j.cub.2005.09.033

Gallup, A. C., & Gallup, G. G. (2008). Yawning and thermoregulation. *Physiology & Behavior*, *95*(1–2), 10–16. https://doi.org/10.1016/J.PHYSBEH.2008.05.003

Giganti, F., & Esposito Ziello, M. (2009). Contagious and spontaneous yawning in autistic and typically developing children. *Current Psychology Letters. Behaviour Brain & Cognition*, *25*(1), 1–12. Retrieved from https://journals.openedition.org/cpl/4810

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., Van De Vorst, M., Van Bon, B. W. M., Willemsen, M. H., … Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, *511*(7509), 344–347. https://doi.org/10.1038/nature13394

Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(3), 173–179. https://doi.org/10.1002/mrdd.20068

Guggisberg, A. G., Mathis, J., Schnider, A., & Hess, C. W. (2010). Why do we yawn? *Neuroscience and Biobehavioral Reviews*, *34*(8), 1267–1276. https://doi.org/10.1016/j.neubiorev.2010.03.008

Guilbaud, O. (2007). L'alexithymie dans ses rapports avec un mode de fonctionnement autistique. *Psychiatrie de l'Enfant*, *50*(2), 503–526. https://doi.org/10.3917/psye.502.0503

Hadjikhani, N., Joseph, R. M., Manoach, D. S., Naik, P., Snyder, J., Dominick, K., … de Gelder, B. (2009). Body expressions of emotion do not trigger fear contagion in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *4*(1), 70–78. https://doi.org/10.1093/scan/nsn038

Hadjikhani, N., Zürcher, N. R., Rogier, O., Hippolyte, L., Lemonnier, E., Ruest, T., … Gillberg, C. (2014). Emotional contagion for pain is intact in autism spectrum disorders. *Translational Psychiatry*, *4*(1), e343–e343. https://doi.org/10.1038/tp.2013.113

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, *26*(6), 822–830.

https://doi.org/doi.org/10.1093/bioinformatics/btq037

Haney, J. L. (2016). Autism, females, and the DSM-5: Gender bias in autism diagnosis. *Social Work in Mental Health*, *14*(4), 396–407. https://doi.org/10.1080/15332985.2015.1031858

Hannant, P., Tavassoli, T., & Cassidy, S. (2016). The role of sensorimotor difficulties in autism spectrum conditions. *Frontiers in Neurology*, *7*(124). https://doi.org/10.3389/fneur.2016.00124

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Harrison, P., & Oakland, T. (2003). *Adaptive Behavior Assessment System – Second Edition*. San Antonio, TX: Harcourt Assessment.

Hasler, B. S., Hirschberger, G., Shani-Sherman, T., & Friedman, D. A. (2014). Virtual peacemakers: Mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior, and Social Networking*, *17*(12), 766–771. https://doi.org/10.1089/cyber.2014.0213

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer.

Hatfield, E. C. (1994). *Emotional Contagion: Studies in Emotion and Social Interaction*. Cambridge: Cambridge University Press.

Hatfield, E. C., Bensman, L., Thornton, P. D., & Rapson, R. L. (2014). *New perspectives on emotional contagion: A review of classic and recent research in facial mimicry and contagion*. *Interpersona: An International Journal on Personal Relationships* (Vol. 8). International Center for Interpersonal Relationship Research. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=5216092

Hatfield, E. C., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–100. https://doi.org/10.1111/1467-8721.ep10770953

Hatfield, E. C., Rapson, R. L., & Yen-Chi, L. Le. (2009). Emotional contagion and empathy. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 19–30). London: MIT Press.

https://doi.org/10.7551/mitpress/9780262012973.003.0015

Haworth, J. L., Libertus, K., & Landa, R. J. (2018). Social intervention impacts action anticipation, goal extraction, and social interest in children with autism spectrum disorder. *Journal of Educational and Developmental Psychology*, *8*(2), 95–106. https://doi.org/10.5539/jedp.v8n2p95

Hein, G., & Singer, T. (2008). I feel how you feel but not always: The empathic brain and its modulation. *Current Opinion in Neurobiology*, *18*(2), 153–158. https://doi.org/10.1016/j.conb.2008.07.012

Heinrich, M., Boehm, J., Sappok, T., Böhm, J., Sappok, T., Boehm, J., & Sappok, T. (2017). Diagnosing autism in adults with intellectual disability: Validation of the DiBAS-R in an independent sample. *Journal of Autism and Developmental Disorders*, *48*(2), 341–350. https://doi.org/10.1007/s10803-017-3336-5

Helt, M. S., Eigsti, I.-M., Snyder, P. J., & Fein, D. A. (2010). Contagious yawning in autistic and typical development. *Child Development*, *81*(5), 1620–1631. https://doi.org/10.1111/j.1467-8624.2010.01495.x

Helt, M. S., & Fein, D. A. (2016). Facial feedback and social input: Effects on laughter and enjoyment in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *46*(1), 83–94. https://doi.org/10.1007/s10803-015-2545-z

Helt, M. S., Fein, D. A., & Vargas, J. E. (2019). Emotional contagion in children with autism spectrum disorder varies with stimulus familiarity and task instructions. *Development and Psychopathology*, *32*(1), 383–393. https://doi.org/10.1017/S0954579419000154

Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. *Journal of Marketing*, *70*(3), 58–73. https://doi.org/10.1509/jmkg.70.3.058

Hertzig, M. E. E., Snow, M. E., & Sherman, M. (1989). Affect and cognition in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*(2), 195–199. https://doi.org/10.1097/00004583-198903000-00008

Hess, U, & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, *40*(2), 129–141. https://doi.org/10.1016/S0167-8760(00)00161-6

Hess, Ursula, & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. https://doi.org/10.1177/1088868312472607

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Hoekstra, R. A., Bartels, M., Verweij, C. J. H., & Boomsma, D. I. (2007). Heritability of autistic traits in the general population. *Archives of Pediatrics and Adolescent Medicine*, *161*(4), 372–377. https://doi.org/10.1001/archpedi.161.4.372

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*(3), 307–316. https://doi.org/10.1037/h0027580

Holle, H., Warne, K., Seth, A. K., Critchley, H. D., & Ward, J. (2012). Neural basis of contagious itch and why some people are more prone to it. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19816–19821. https://doi.org/10.1073/pnas.1216160109

Hoogeveen, A. T., & Oostra, B. A. (1997). The fragile X syndrome. *Journal of Inherited Metabolic Disease*, *20*(2), 139–151. https://doi.org/10.1023/A:1005392319533

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2010). Party: A laboratory for recursive partytioning. https://doi.org/10.1198/106186006X133933

Hudenko, W. J., Stone, W., & Bachorowski, J. A. (2009). Laughter differs in children with autism: An acoustic analysis of laughs produced by children with and without the disorder. *Journal of Autism and Developmental Disorders*, *39*(10), 1392–1400. https://doi.org/10.1007/s10803-009-0752-1

Hull, L., Lai, M. C., Baron-Cohen, S., Allison, C., Smith, P., Petrides, K. V., & Mandy, W. (2020).

Gender differences in self-reported camouflaging in autistic and non-autistic adults. *Autism*, *24*(2), 352–363. https://doi.org/10.1177/1362361319864804

Hull, L., Mandy, W., Lai, M.-C., Baron-Cohen, S., Allison, C., Smith, P., & Petrides, K. V. (2018). Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). *Journal of Autism and Developmental Disorders*, *49*(3), 1–15. https://doi.org/10.1007/s10803-018-3792-6

Humphrey, N., & Lewis, S. (2008). `Make me normal'. *Autism*, *12*(1), 23–46. https://doi.org/10.1177/1362361307085267

Hurlbutt, K., & Chalmers, L. (2002). Adults with autism speak out. *Focus on Autism and Other Developmental Disabilities*, *17*(2), 103–111. https://doi.org/10.1177/10883576020170020501

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the autism spectrum quotient (AQ) in a non-clinical sample. *Personality and Individual Differences*, *43*, 1938–1949. https://doi.org/10.1016/j.paid.2007.06.012

Hurtado, B., Jones, L., & Burniston, F. (2014). Is easy read information really easier to read? *Journal of Intellectual Disability Research*, *58*(9), 822–829. https://doi.org/10.1111/jir.12097

ICM. (2016). *British Medical Association survey of GPs in England, October–November 2016*. Retrieved from www.bma.org.uk/collective-voice/influence/key-negotiations/training-and-workforce/urgent-prescription-forgeneral-practice/key-issues-survey#Workload

Ip, A., Zwaigenbaum, L., & Brian, J. A. (2019). Post-diagnostic management and follow-up care for autism spectrum disorder. *Paediatrics & Child Health*, *24*(7), 461–468. https://doi.org/10.1093/PCH/PXZ121

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning - with Applications in R*. Retrieved from https://www.springer.com/gp/book/9781461471370%0Ahttp://www.springer.com/us/book/9781461471370

Janitza, S., Strobl, C., & Boulesteix, A. L. (2013). An AUC-based permutation variable importance

measure for random forests. *BMC Bioinformatics*, *14*(1), 119. https://doi.org/10.1186/1471-2105-14-119

Javaid, A., Nakata, V., & Michael, D. (2019). Diagnostic overshadowing in learning disability: Think beyond the disability. *Progress in Neurology and Psychiatry*, *23*(2), 8–10. https://doi.org/10.1002/pnp.531

Jiménez, L., Lorda, M. J., & Méndez, C. (2014). Emulation and mimicry in school students with typical development and with high functioning autism. *Journal of Autism and Developmental Disorders*, *44*(7), 1597–1608. https://doi.org/10.1007/s10803-013-2027-0

Jones, A. P., Happé, F., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: Different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *51*(11), 1188–1197. https://doi.org/10.1111/j.1469-7610.2010.02280.x

Jones, L., Goddard, L., Hill, E. L., Henry, L. A., & Crane, L. (2014). Experiences of receiving a diagnosis of autism spectrum disorder: A survey of adults in the United Kingdom. *Journal of Autism and Developmental Disorders*, *44*, 3033–3044. https://doi.org/10.1007/s10803-014-2161-3

Jordan, E. M., & Thomas, D. G. (2017). Contagious positive affective responses to laughter in infancy. *Archives of Psychology*, *1*(2). Retrieved from https://archivesofpsychology.org/index.php/aop/article/view/18

Jung, W. E. (2003). The inner eye theory of laughter: Mindreader signals cooperator value. *Evolutionary Psychology*, *1*(1). https://doi.org/10.1177/147470490300100118

Juslin, P. N., Harmat, L., & Laukka, P. (2018). The wisdom of the body: Listeners' autonomic arousal distinguishes between spontaneous and posed vocal emotions. *Scandinavian Journal of Psychology*, *59*(2), 105–112. https://doi.org/10.1111/sjop.12429

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kenny, H., & Stansfield, A. J. (2016). How useful are the adult asperger assessment and AQ-10 within

an adult clinical population of all intellectual abilities? *Advances in Autism*, *2*(3), 118–130. https://doi.org/10.1108/AIA-03-2016-0009

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, *20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., … Vincent, J. B. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *American Journal of Human Genetics*, *90*(5), 856–863. https://doi.org/10.1016/j.ajhg.2012.03.023

Kline, P. (1998). *The New Psychometrics* (First). London: Routledge.

Kloosterman, P. H., Keefer, K. V, Kelley, E. A., Summerfeldt, L. J., & Parker, J. D. A. (2011). Evaluation of the factor structure of the autism-spectrum quotient. *Personality and Individual Differences*, *50*, 310–314. https://doi.org/10.1016/j.paid.2010.10.015

Koegel, L. K., Koegel, R. L., Ashbaugh, K., & Bradshaw, J. (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. *International Journal of Speech-Language Pathology*, *16*(1), 50–56. https://doi.org/10.3109/17549507.2013.861511

Koudys, J., Perry, A., Ho, H., & Charles, M. (2018). Psychosocial outcomes in adoescents with autism who received intensive behavioural intervention as young children. *Journal on Developmental Disabilities*, *23*(2), 85.

Kraijer, D. (1990). *Autisme- en Verwante kontakt- stoornissenschaal voor Zwakzinnigen. Handleiding (Autism and related contact disorders scale for the mentally retarded. Manual)*. Lisse, The Netherlands: Swets and Zeitlinger.

Kraijer, D., & de Bildt, A. (2005). The PDD-MRS: An instrument for identification of autism spectrum disorders in persons with mental retardation. *Journal of Autism and Developmental Disorders*, *35*(4), 499–513. https://doi.org/http://dx.doi.org/10.1007/s10803-005-5040-0

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kringelbach, M. L., Stark, E. A., Alexander, C., Bornstein, M. H., & Stein, A. (2016). On cuteness: Unlocking the parental brain and beyond. *Trends in Cognitive Sciences*, *20*(7), 545–558. https://doi.org/10.1016/j.tics.2016.05.003

Krug, D. A., Arick, J., & Almond, P. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry*, *21*(3), 221–229. https://doi.org/10.1111/j.1469-7610.1980.tb01797.x

La Malfa, G., Lassi, S., Bertelli, M., Salvini, R., & Placidi, G. F. (2004). Autism and intellectual disability: A study of prevalence on a sample of the Italian population. *Journal of Intellectual Disability Research*, *48*(3), 262–267. https://doi.org/10.1111/j.1365-2788.2003.00567.x

Lai, M. C., Lombardo, M. V., Ruigrok, A. N. V., Chakrabarti, B., Auyeung, B., Szatmari, P., … Baron-Cohen, S. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, *21*(6), 690–702. https://doi.org/10.1177/1362361316671012

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42–58. https://doi.org/10.1162/jocn.2007.19.1.42

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 374. https://doi.org/10.2307/2529786

Langridge, A. T., Glasson, E. J., Nassar, N., Jacoby, P., Pennell, C., Hagan, R., … Stanley, F. J. (2013). Maternal conditions and perinatal characteristics associated with autism spectrum disorder and intellectual disability. *PLOS ONE*, *8*(1), e50963. https://doi.org/10.1371/journal.pone.0050963

Lau, W. Y. P., Kelly, A. B., & Peterson, C. C. (2013). Further evidence on the factorial structure of the autism spectrum quotient (AQ) for adults with and without a clinical diagnosis of autism. *Journal*

of Autism and Developmntal Disorders, *43*(12), 2807–2815. https://doi.org/10.1007/s10803-013-1827-6

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*(3), 301–310. https://doi.org/10.1023/B:JADD.0000029552.42724.1b

Leask, J., Leask, A., & Silove, N. (2005). Evidence for autism in folklore? *Archives of Disease in Childhood*, *90*(3), 271. https://doi.org/10.1136/adc.2003.044958

Lesser, I. M. (1981). A review of the alexithymia concept. *Psychosomatic Medicine*. https://doi.org/10.1097/00006842-198112000-00009

Leung, K., Trevena, L., & Waters, D. (2012). Development of an appraisal tool to evaluate strength of an instrument or outcome measure. *Nurse Researcher*, *20*(2), 13–19. https://doi.org/10.7748/nr2012.11.20.2.13.c9436

Levy, S. G., & Fenley, W. F. (1979). Audience size and likelihood and intensity of response during a humorous movie. *Bulletin of the Psychonomic Society*, *13*(6), 409–412.

Li, C., Zhou, H., Wang, T., Long, S., Du, X., Xu, X., … Wang, Y. (2018). Performance of the autism spectrum rating scale and social responsiveness scale in identifying autism spectrum disorder among cases of intellectual disability. *Neuroscience Bulletin*, *34*(6), 972–980. https://doi.org/10.1007/s12264-018-0237-3

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. Retrieved from http://www.stat.berkeley.edu/

Licari, M. K., Alvares, G. A., Varcin, K., Evans, K. L., Cleary, D., Reid, S. L., … Whitehouse, A. J. O. (2020). Prevalence of motor difficulties in autism spectrum disorder: Analysis of a population-based cohort. *Autism Research*, *13*(2), 298–306. https://doi.org/10.1002/aur.2230

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., & Weyers, P. (2011). Processes underlying congruent and incongruent facial reactions to emotional facial expressions. *Emotion*, *11*(3), 457–

467. https://doi.org/10.1037/a0023162

Lindsay, W., Neilson, C., & Lawrenson, H. (1997). Cognitive-Behaviour Therapy For People With learning Disabilities. In B. S. Kroese, D. Dagnan, & K. Loumidis (Eds.), *Cognitive-Behaviour Therapy for People with Learning Disabilities*. London: Routledge.

Lippold, T., & Burns, J. (2009). Social support and intellectual disabilities: A comparison between social networks of adults with intellectual disability and those with physical disability. *Journal of Intellectual Disability Research*, *53*(5), 463–473. https://doi.org/10.1111/j.1365-2788.2009.01170.x

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(6), 466. https://doi.org/10.1016/j.jaac.2017.03.013

Lopez-Raton, M., Rodriguez-Alvarez, M. X., Suarez, C. C., & Gude Sampedro, F. (2014). Optimal cutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, *61*(8), 1. Retrieved from http://www.biosoft.hacettepe.edu.tr/easyROC/

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11055457

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Loveland, K. A., Tunali-Kotoski, B., Pearson, D. A., Brelsford, K. A., Ortegon, J., & Chen, R. (1994). Imitation and expression of facial affect in autism. *Development and Psychopathology*, *6*(3), 433–444. https://doi.org/10.1017/S0954579400006039

Lovell, B., Moss, M., & Wetherell, M. (2012). The psychosocial, endocrine and immune consequences

of caring for a child with autism or ADHD. *Psychoneuroendocrinology*, *37*(4), 534–542. https://doi.org/10.1016/j.psyneuen.2011.08.003

Lu, J. S., Chen, Q. Y., Zhou, S. B., Wu, F. Y., Liu, R. H., Zhou, Z. X., … Zhuo, M. (2019). Contagious itch can be induced in humans but not in rodents. *Molecular Brain*, *12*(1), 38. https://doi.org/10.1186/s13041-019-0455-2

Lunsky, Y. (2006). Individual differences in interpersonal relationships for persons with mental retardation. *International Review of Research in Mental Retardation*, *31*, 117–161. https://doi.org/10.1016/S0074-7750(05)31004-4

Magyar, C. I., Pandolfi, V., & Dill, C. A. (2012). An initial evaluation of the social communication questionnaire for the assessment of autism spectrum disorders in children with down syndrome. *Journal of Developmental and Behavioral Pediatrics*, *33*(2), 134. https://doi.org/10.1097/DBP.0b013e318240d3d9

Maloret, P., & Sumner, K. (2014). Understanding autism spectrum conditions. *Learning Disability Practice*, *17*(6), 23–26. https://doi.org/10.7748/ldp.17.6.23.e1537

Malow, B. A., Katz, T., Reynolds, A. M., Shui, A., Carno, M., Connolly, H. V., … Bennett MD, MPH, A. E. (2016). Sleep difficulties and medications in children with autism spectrum disorders: A registry study. *Pediatrics*, *137*(2), S98–S104. https://doi.org/10.1542/peds.2015-2851H

Mann, J., McDermott, S., Hardin, J., Pan, C., & Zhang, Z. (2013). Pre-pregnancy body mass index, weight change during pregnancy, and risk of intellectual disability in children. *BJOG: An International Journal of Obstetrics & Gynaecology*, *120*(3), 309–319. https://doi.org/10.1111/1471-0528.12052

Markram, K., & Markram, H. (2010). The intense world theory - A unifying theory of the neurobiology of autism. *Frontiers in Human Neuroscience*, *4*, 224. https://doi.org/10.3389/fnhum.2010.00224

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Massen, J. J. M., & Gallup, A. C. (2017). Why contagious yawning does not (yet) equate to empathy. *Neuroscience and Biobehavioral Reviews*, *80*, 573–585. https://doi.org/10.1016/j.neubiorev.2017.07.006

Massey, L., & Williams, S. (2006). Implementing change: The perspective of NHS change agents. *Leadership and Organization Development Journal*, *27*(8), 667–681. https://doi.org/10.1108/01437730610709282

Matson, J. L. (1995). *The Matson evaluation of social skills for individuals with severe mental retardation*. Baton Rouge, LA: Scientific Publishers.

Matson, J. L., Fodstad, J. C., & Mahan, S. (2009). Cutoffs, norms, and patterns of comorbid difficulties in children with developmental disabilities on the baby and infant screen for children with autism traits (BISCUIT-Part 2). *Research in Developmental Disabilities*, *30*(6), 1221–1228. https://doi.org/10.1016/j.ridd.2009.04.004

Matson, J. L., & Minshawi, N. F. (2006). *Early Intervention For Autism Spectrum Disorders: A Critical Analysis* (1st ed.). Oxford: Elsevier.

Matson, J. L., & Rivet, T. T. (2007). The effects of severity of autism and PDD-NOS symptoms on challenging behaviors in adults with intellectual disabilities. *Journal of Developmental and Physical Disabilities*, *20*(1), 41–51. https://doi.org/10.1007/s10882-007-9078-0

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*, *30*(6), 1107–1114. https://doi.org/10.1016/J.RIDD.2009.06.003

Matson, J. L., Wilkins, J., Boisjoli, J. A., & Smith, K. R. (2008). The validity of the autism spectrum disorders-diagnosis for intellectually disabled adults (ASD-DA). *Research in Developmental Disabilities*, *29*(6), 537–546. https://doi.org/10.1016/J.RIDD.2007.09.006

Matson, J. L., Wilkins, J., & González, M. (2007). Reliability and factor structure of the autism spectrum disorders— Diagnosis scale for intellectually disabled adults (ASD—DA). *Journal of Developmental and Physical Disabilities*, *19*(6), 565–577. https://doi.org/10.1007/s10882-007-

9070-8

Mcclure, K. S., Halpern, J., Wolper, P. A., & Donahue, J. (2009). Emotion regulation and intellectual disability. *Journal on Developmental Disabilities*, *15*(2), 38–44. Retrieved from https://mdsoar.org/handle/11603/4016

McDermott, S., Bao, W., Tong, X., Cai, B., Lawson, A., & Aelion, C. M. (2014). Are different soil metals near the homes of pregnant women associated with mild and severe intellectual disability in children? *Developmental Medicine & Child Neurology*, *56*(9), 888–897. https://doi.org/10.1111/dmcn.12442

Mcdermott, S., Mann, J. R., & Hardin, J. (2015). Preventing intellectual disability during pregnancy: What are the potentially high yield targets? *Article in Journal of Perinatal Medicine*. https://doi.org/10.1515/jpm-2015-0059

McDonnell, C. G., DeLucia, E. A., Hayden, E. P., Anagnostou, E., Nicolson, R., Kelley, E., … Stevenson, R. A. (2019). An exploratory analysis of predictors of youth suicide-related behaviors in autism spectrum disorder: Implications for prevention science. *Journal of Autism and Developmental Disorders*, 1–14. https://doi.org/10.1007/s10803-019-04320-6

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23092060

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science*, *9*(3), 295–302. https://doi.org/10.1111/j.1467-7687.2006.00492.x

McKenzie, K., Forsyth, K., O'Hare, A., McClure, I., Rutherford, M., Murray, A., & Irvine, L. (2015). Factors influencing waiting times for diagnosis of autism spectrum disorder in children and adults. *Research in Developmental Disabilities*, *45–46*, 300–306. https://doi.org/10.1016/j.ridd.2015.07.033

McKenzie, K., & Murray, A. L. (2015). Evaluating the use of the child and adolescent intellectual disability screening questionnaire (CAIDS-Q) to estimate IQ in children with low intellectual

ability. *Research in Developmental Disabilities*, *37*, 31–36. https://doi.org/10.1016/j.ridd.2014.11.001

McNeill, J. (2019). Social validity and teachers' use of evidence-based practices for autism. *Journal of Autism and Developmental Disorders*, *49*(11), 4585–4594. https://doi.org/10.1007/s10803-019-04190-y

Mefford, H. C., Batshaw, M. L., & Hoffman, E. P. (2012). Genomics, intellectual disability, and autism. *Genomic Medicine*, *366*(8), 733–743. https://doi.org/doi:10.1056/NEJMra1114194

Metcalfe, C., Lewin, R., Wisher, S., Perry, S., Bannigan, K., & Moffett, J. K. (2001). Barriers to implementing the evidence base in four NHS therapies. *Physiotherapy*, *87*(8), 433–441. https://doi.org/10.1016/S0031-9406(05)65462-4

Mickan, S. M. (2005). Evaluating the effectiveness of health care teams. *Australian Health Review*, *29*(2), 211–217. https://doi.org/10.1071/AH050211

Microsoft. (2020). Microsoft Azure: Facial Recognition. Face. An AI service that analyses faces in images. Retrieved August 24, 2020, from https://azure.microsoft.com/en-gb/services/cognitive-services/face/

Midence, K., & O'Neill, M. (1999). The experience of parents in the diagnosis of autism: A pilot study. *Autism*, *3*(3), 273–285. https://doi.org/10.1177/1362361399003003005

Mockett, M., Khan, J., & Theodosiou. (2011). Parental perceptions of a Manchester service for autistic spectrum disorders. *International Journal of Family Medicine*, 1–6. Retrieved from https://www.hindawi.com/journals/ijfm/2011/601979/

Moh, T. A., & Magiati, I. (2012). Factors associated with parental stress and satisfaction during the process of diagnosis of children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(1), 293–303. https://doi.org/10.1016/J.RASD.2011.05.011

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion*, *7*(2), 447–457.

https://doi.org/10.1037/1528-3542.7.2.447

Murphy, G. (2017). The NICE guidelines and quality standards on learning disabilities and behaviour that challenges. *Tizard Learning Disability Review*, *22*(2), 71–81. https://doi.org/10.1108/TLDR-12-2016-0044

Murray, A. L., Booth, T., McKenzie, K., & Kuenssberg, R. (2015). What range of trait levels can the autism-spectrum quotient (AQ) measure reliably? An item response theory analysis. *Psychological Assessment*, *28*(6), Advance online publication. https://doi.org/http://dx.doi.org/10.1037/pas0000215

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the autism spectrum quotient short form. *Journal of Autism and Developmental Disorders*, *44*(1), 55–64. https://doi.org/10.1007/s10803-013-1851-6

Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2015). Do the autism spectrum quotient (AQ) and autism spectrum quotient short form (AQ-S) Primarily reflect general ASD traits or specific ASD traits? A bi-factor analysis. *Assessment*, 1–14. https://doi.org/10.1177/1073191115611230

Mutsaerts, C. G., Heinrich, M., Sterkenburg, P. S., & Sappok, T. (2016). Screening for ASD in adults with ID-moving toward a standard using the DiBAS-R and the ACL. *Journal of Intellectual Disability Research*, *60*(5), 512–522. https://doi.org/http://dx.doi.org/10.1111/jir.12290

National Autism Resources. (2020). The Picture Exchange Communication System (PECS). Retrieved July 9, 2020, from https://www.nationalautismresources.com/the-picture-exchange-communication-system-pecs/

National Autistic Society. (2020). National Autistic Society - Autism support - Leading UK charity. Retrieved July 21, 2020, from https://www.autism.org.uk/

National Institute for Health and Care Excellence (NICE). (2016). Autism spectrum disorder in adults: Diagnosis and management. Retrieved from https://www.nice.org.uk/guidance/CG142/chapter/1-

Guidance#interventions-for-autism

National Institute for Health and Care Excellence (NICE). (2018, October). Learning disabilities: When should I suspect a learning disability? Retrieved May 12, 2020, from https://cks.nice.org.uk/learning-disabilities#!diagnosisSub

National Institute for Health and Care Excellence (NICE), & NHS England. (2016). *Proposals for changes to the arrangements for evaluating and funding drugs and other health technologies appraised through NICE's technology appraisal and highly specialised technologies programmes*. Retrieved from https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/technology-appraisals/NICE_NHSE_TA_and_HST_consultation_document.pdf

National Institute for Health and Clinical Excellence (NICE). (2011). NICE Clinical Guideline. Autism: Recognition, referral and diagnosis of children and young people on the autism spectrum. *National Collaborating Centre for Women's and Children's Health*, *85*(1), 22–24.

National Institute for Health and Clinical Excellence [NICE]. (2012). *Autism: Recognition, referral, diagnosis and management of adults on the autism spectrum [CG23]*. London: NICE.

National Institute for Health and Clinical Excellence [NICE]. (2015). Challenging behaviour and learning disabilities: Prevention and interventions for people with learning disabilities whose behaviour challenges. NG11., (May).

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678. https://doi.org/10.1177/1948550611406138

Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: SAGE Publications.

Neves, L., Cordeiro, C., Scott, S. K., Castro, S. L., & Lima, C. F. (2018). High emotional contagion and empathy are associated with enhanced detection of emotional authenticity in laughter. *Quarterly Journal of Experimental Psychology*, *71*(11), 2355–2363. https://doi.org/10.1177/1747021817741800

NICE. (2012). Autism spectrum disorder in adults: Diagnosis and management. *National Institute of Clinical Excellence*. Retrieved from https://www.nice.org.uk/guidance/cg142/chapter/1-guidance

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

Niemeier, V., & Gieler, U. (2000). Observations during itch-inducing lecture. *Dermatology and Psychosomatics / Dermatologie Und Psychosomatik*, *1*(1), 15–18. https://doi.org/10.1159/000057993

Noens, I., & van Berckelaer-Onnes, I. (2004). Making sense in a fragmentary world: Communication in people with autism and learning disability. *Autism: The International Journal of Research and Practice*, *8*(2), 197–218. https://doi.org/10.1177/1362361304042723

Nota, L., Ferrari, L., Soresi, S., & Wehmeyer, & M. (2007). Self-determination, social abilities and the quality of life of people with intellectual disability. *Journal of Intellectual Disability Research*, *51*(11), 850–865. https://doi.org/10.1111/j.1365-2788.2006.00939.x

O'Brien, G., & Pearson, J. (2004). Autism and learning disability, *8*(2), 125–140. https://doi.org/10.1177/1362361304042718

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, *32*(3), 396–402. https://doi.org/10.3758/bf03200807

O'Haire, M. E. (2017, July 3). Research on animal-assisted intervention and autism spectrum disorder, 2012–2015. *Applied Developmental Science*. Psychology Press. https://doi.org/10.1080/10888691.2016.1243988

O'Haire, M. E., McKenzie, S. J., Beck, A. M., & Slaughter, V. (2013). Social behaviors increase in children with autism in the presence of animals compared to toys. *PLoS ONE*, *8*(2), e57010. https://doi.org/10.1371/journal.pone.0057010

O'Haire, M. E., McKenzie, S. J., Beck, A. M., & Slaughter, V. (2015). Animals may act as social

buffers: Skin conductance arousal in children with autism spectrum disorder in a social context. *Developmental Psychobiology*, *57*(5), 584–595. https://doi.org/10.1002/dev.21310

O'Leary, C., Leonard, H., Bourke, J., D'Antoine, H., Bartu, A., & Bower, C. (2013). Intellectual disability: Population-based estimates of the proportion attributable to maternal alcohol use disorder during pregnancy. *Developmental Medicine & Child Neurology*, *55*(3), 271–277. https://doi.org/10.1111/dmcn.12029

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2009). Slow echo: Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Developmental Science*, *12*(4), 510–520. https://doi.org/10.1111/j.1467-7687.2008.00796.x

OpenCV. (2020). Open Source Computer Vision (OpenCV) Library. Retrieved August 28, 2020, from https://opencv.org/

Osborne, L. A., & Reed, P. (2008). Parents' perceptions of communication with professionals during the diagnosis of autism. *Autism*, *12*(3), 309–324. https://doi.org/10.1177/1362361307089517

Oswald, D. P., & Volkmar, F. R. (1991). Signal detection analysis of items from the autism behavior checklist. *Journal of Autism and Developmental Disorders*, *21*(Dec 91), 543–549. https://doi.org/10.1007/BF02206876

Palagi, E., Leone, A., Mancini, G., & Ferrari, P. F. (2009). Contagious yawning in gelada baboons as a possible expression of empathy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19262–19267. https://doi.org/10.1073/pnas.0910891106

Pandolfi, V., Magyar, C. I., & Dill, C. A. (2018). Screening for autism spectrum disorder in children with down syndrome: An evaluation of the pervasive developmental disorder in mental retardation scale. *Journal of Intellectual & Developmental Disability*, *43*(1), 61–72. https://doi.org/10.3109/13668250.2016.1271111

Papoiu, A. D. P., Wang, H., Coghill, R. C., Chan, Y.-H., & Yosipovitch, G. (2011). Contagious itch in humans: A study of visual 'transmission' of itch in atopic dermatitis and healthy subjects. *British*

Journal of Dermatology, *164*(6), 1299–1303. https://doi.org/10.1111/j.1365-2133.2011.10318.x

Pathak, V., Jena, B., & Kalra, S. (2013). Qualitative research. *Perspectives in Clinical Research*, *4*(3). https://doi.org/10.4103/2229-3485.115389

Patterson, D. (1987). The causes of down syndrome. *Scientific American*, *257*(2), 52–61. https://doi.org/10.2307/24979443

Paukner, A., Suomi, S. J., Visalberghi, E., & Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science (New York, N.Y.)*, *325*(5942), 880–883. https://doi.org/10.1126/science.1176269

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–261. https://doi.org/10.1023/A:1016374617369

Peterson, B. L. (2017). Thematic Analysis/Interpretive Thematic Analysis. In *The International Encyclopedia of Communication Research Methods* (pp. 1–9). Wiley. https://doi.org/10.1002/9781118901731.iecrm0249

Platek, S. M., Critton, S. R., Myers, T. E., Gallup Jr., G. G., & Gallup, G. G. (2003). Contagious yawning: The role of self-awareness and mental state attribution. *Cognitive Brain Research*, *17*(2), 223–227. https://doi.org/10.1016/S0926-6410(03)00109-5

Platek, S. M., Mohamed, F. B., & Gallup, G. G. (2005). Contagious yawning and the brain. *Cognitive Brain Research*, *23*(2–3), 448–452. https://doi.org/10.1016/J.COGBRAINRES.2004.11.011

Poncelas, A., & Murphy, G. (2007). Accessible information for people with intellectual disabilities: Do symbols really help? *Journal of Applied Research in Intellectual Disabilities*, *20*(5), 466–474. https://doi.org/10.1111/j.1468-3148.2006.00334.x

Poquérusse, J., Pastore, L., Dellantonio, S., & Esposito, G. (2018). Alexithymia and autism spectrum disorder: A complex relationship. *Frontiers in Psychology*, *9*(JUL), 1196. https://doi.org/10.3389/fpsyg.2018.01196

Preston, S. D., & de Waal, F. B. M. (2001). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/S0140525X02000018

Provine, R. R. (1986). Yawning as a stereotyped action pattern and releasing stimulus. *Ethology*, *72*(2), 109–122. https://doi.org/10.1111/j.1439-0310.1986.tb00611.x

Provine, R. R. (1989). Faces as releasers of contagious yawning: An approach to face detection using normal human subjects. *Bulletin of the Psychonomic Society*, *27*(3), 211–214. https://doi.org/10.3758/BF03334587

Provine, R. R. (1992). Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bulletin of the Psychonomic Society*, *30*(1), 1–4. https://doi.org/10.3758/BF03330380

Provine, R. R. (1996). Contagious Yawning and Laughter: Significance for Sensory Feature Detection, Motor Pattern Generation, Imitation, and the Evolution of Social Behaviour. In C. M. Heyes & B. G. Galef (Eds.), *Social Learning In Animals: The Roots of Culture* (pp. 179–210).

Provine, R. R. (2004). Laughing, tickling, and the evolution of speech and self. *Current Directions in Psychological Science*, *13*(6), 215–218. https://doi.org/10.1111/j.0963-7214.2004.00311.x

Provine, R. R. (2005). Yawning: The yawn is primal, unstoppable and contagious, revealing the evolutionary and neural basis of empathy and unconscious behavior. *American Scientist*. Sigma Xi, The Scientific Research Honor Society. https://doi.org/10.2307/27858677

Provine, R. R., & Emmorey, K. (2006). Laughter among deaf signers. *Journal of Deaf Studies and Deaf Education*, *11*(4), 403. https://doi.org/10.1093/DEAFED/ENL008

Public Health England. (2019). NHS population screening explained - GOV.UK. Retrieved March 9, 2020, from https://www.gov.uk/guidance/nhs-population-screening-explained

Pugh, S. D. (2001). Service with a smile: Emotional contagion in the service encounter. *Academy of*

Management Journal, *44*(5), 1018–1027. https://doi.org/10.5465/3069445

R Core Team. (2013). *R: A language and environment for statistical computing* (Vienna, Au). R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rao, P. A., & Beidel, D. C. (2009). The impact of children with high-functioning autism on parental stress, sibling adjustment, and family functioning. *Behavior Modification*, *33*(4), 437–451. https://doi.org/10.1177/0145445509336427

Raykov, T. (2002). Examining group differences in reliability of multiple-component instruments. *British Journal of Mathematical and Statistical Psychology*, *55*(1), 145–158. https://doi.org/10.1348/000711002159743

Reddy, V., Williams, E., & Vaughan, A. (2002). Sharing humour and laughter in autism and down's syndrome. *British Journal of Psychology*, *93*, 219–242. https://doi.org/10.1348/000712602162553

Reilly, C. (2009). Autism spectrum disorders in down syndrome: A review. *Research in Autism Spectrum Disorders*, *3*(4), 829–839. https://doi.org/10.1016/j.rasd.2009.01.012

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. a. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Richardson, S., Mishuris, R., O'Connell, A., Feldstein, D., Hess, R., Smith, P., … Mann, D. (2017). "Think aloud" and "near live" usability testing of two complex clinical decision support tools. *International Journal of Medical Informatics*, *106*, 1–8. https://doi.org/10.1016/j.ijmedinf.2017.06.003

Rivard, M., Forget, J., Giroux, N., Mello, C., Kerr, K., & Regli, G. (2016). Observation of socially appropriate and inappropriate behaviours among children with autism spectrum disorder during an early behavioural intervention program. *Journal on Developmental Disabilities*, *22*(1), 52–67.

Robertson, J., Hatton, C., Felce, D., Meek, A., Carr, D., Knapp, M., … Lowe, K. (2005). Staff stress and morale in community-based settings for people with intellectual disabilities and challenging

behaviour: A brief report. *Journal of Applied Research in Intellectual Disabilities*, *18*(3), 271–277. https://doi.org/10.1111/j.1468-3148.2005.00233.x

Robins, D. L., Fein, D., & Barton, M. (2009). *Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F)*. Retrieved from http://www.mchatscreen.com.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–144. https://doi.org/10.1023/A:1010738829569

Rodgers, J. (1999). An examination of the imitation deficit in autism. In J. Nadel & G. Butterworth (Eds.), *Cambridge studies in cognitive perceptual development.* (pp. 254–283). New York: Cambridge University Press. Retrieved from http://psycnet.apa.org/record/1999-02466-009

Rogers, C. L., Goddard, L., Hill, E. L., Henry, L. A., & Crane, L. (2016). Experiences of diagnosing autism spectrum disorder: A survey of professionals in the United Kingdom. *Autism*, *20*(7), 820–831.

Rooney, J., Harris, E., & Collins, H. (2018). An evaluation of the health equalities framework for people with a learning disability. *Learning Disability Practice*, *21*, 32–37. https://doi.org/10.7748/ldp.2018.e1886

Royal College of Nursing. (2016). *RCN Nurse Survey*. Retrieved from www.rcn.org.uk/-/media/%0Aroyal-college-of-nursing/documents/forums/children-and-young-people/staying-healthy-forum/rcn-surveyof-school-nurses-2016.pdf

Ruiz Calzada, L., Pistrang, N., & Mandy, W. P. L. (2012). High-functioning autism and asperger's disorder: Utility and meaning for families. *Journal of Autism and Developmental Disorders*, *42*, 230–243. https://doi.org/10.1007/s10803-011-1238-5

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*(4), 585–615. https://doi.org/10.1037/0096-3445.112.4.585

Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019, March 1). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. *Molecular Autism*. BioMed Central Ltd. https://doi.org/10.1186/s13229-019-0260-x

Ruzich, E., Allison, C., Chakrabarti, B., Smith, P., Musto, H., Ring, H., & Baron-Cohen, S. (2015). Sex and STEM occupation predict autism-spectrum quotient (AQ) scores in half a million people. *PLOS ONE*, *10*(10), e0141229. https://doi.org/10.1371/journal.pone.0141229

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the autism-spectrum quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 2. https://doi.org/10.1186/2040-2392-6-2

Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K., & Szatkowska, I. (2016). Emotional empathy and facial mimicry for static and dynamic facial expressions of fear and disgust. *Frontiers in Psychology*, *7*(1853), 1–11. https://doi.org/10.3389/fpsyg.2016.01853

Rynkiewicz, A., Schuller, B., Marchi, E., Piana, S., Camurri, A., Lassalle, A., & Baron-Cohen, S. (2016). An investigation of the "female camouflage effect" in autism using a computerized ADOS-2 and a test of sex/gender differences. *Molecular Autism*, *7*(1), 1–8. https://doi.org/10.1186/s13229-016-0073-0

Saemundsen, E., Juliusson, H., Hjaltested, S., Gunnarsdottir, T., Halldorsdottir, T., Hreidarsson, S., & Magnusson, P. (2010). Prevalence of autism in an urban population of adults with severe intellectual disabilities - a preliminary study. *Journal of Intellectual Disability Research*, *54*(8), 727–735. https://doi.org/10.1111/j.1365-2788.2010.01300.x

Saha, A., & Agarwal, N. (2016). Modeling social support in autism community on social media. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *5*(8), 1–14. https://doi.org/10.1007/s13721-016-0115-8

Saletta, M., Kaldenberg, E., Rivera, K., & Wood, A. (2019). Do illustrations promote reading comprehension in adults with intellectual or developmental disabilities? Education and training in

autism and developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, *54*(3), 225–236. Retrieved from https://eric.ed.gov/?id=EJ1226032

Sams, M. J., Fortney, E. V., & Willenbring, S. (2006, May 1). Occupational therapy incorporating animals for children with autism: A pilot investigation. *American Journal of Occupational Therapy*. American Occupational Therapy Association, Inc. https://doi.org/10.5014/ajot.60.3.268

Sappok, T., Brooks, W., Heinrich, M., McCarthy, J., & Underwood, L. (2017). Cross-cultural validity of the social communication questionnaire for adults with intellectual developmental disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 393–404. https://doi.org/10.1007/s10803-016-2967-2

Sappok, T., Diefenbacher, A., Gaul, I., & Bölte, S. (2015). Validity of the social communication questionnaire in adults with intellectual disabilities and suspected autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, *120*(3), 203–214.

Sappok, T., Gaul, I., Bergmann, T., Dziobek, I., Bölte, S., Diefenbacher, A., & Heinrich, M. (2014). The diagnostic behavioral assessment for autism spectrum disorder—revised: A screening instrument for adults with intellectual disability suspected of autism spectrum disorders. *Research in Autism Spectrum Disorders*, *8*(4), 362–375. https://doi.org/10.1016/J.RASD.2013.12.016

Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, *104*(1), 1–18. https://doi.org/10.1016/j.cognition.2006.05.001

Scambler, D. J., Hepburn, S., Rutherford, M. D., Wehner, E. A., & Rogers, S. J. (2007). Emotional responsivity in children with autism, children with other developmental disabilities, and children with typical development. *Journal of Autism and Developmental Disorders*, *37*(3), 553–563. https://doi.org/10.1007/s10803-006-0186-y

Scheub, R., Sorenson, T., & Helt, M. (2020). The role of empathy, clinical traits, and eye gaze in contagious yawning and itching. In *Trinity College, Hartford*. Retrieved from https://commons.trincoll.edu/brain/files/2020/05/Rachel-Scheub-2020-symposium.pdf

Schineller, M. (2018). Increased contagious itch in children with autism spectrum disorder (ASD).

Senior Theses and Projects. Retrieved from https://digitalrepository.trincoll.edu/theses/705

Schürmanna, M., Hesse, M. D., Stephan, K. E., Saarela, M., Zilles, K., Hari, R., & Fink, G. R. (2005). Yearning to yawn: The neural basis of contagious yawning. *NeuroImage*, *24*(4), 1260–1264. https://doi.org/10.1016/J.NEUROIMAGE.2004.10.022

Scotland, J. L., McKenzie, K., Cossar, J., Murray, A., Michie, A., & McKenzie, K. (2016). The ability of adults with an intellectual disability to recognise facial expressions of emotion in comparison with typically developing individuals: A systematic review. *Research in Developmental Disabilities*, *41–42*, 69–78. https://doi.org/10.1016/j.ridd.2015.05.007

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, *18*(12), 618–620. https://doi.org/10.1016/j.tics.2014.09.002

Seibt, B., Mühlberger, A., Likowski, K. U., & Weyers, P. (2015). Facial mimicry in its social setting. *Frontiers in Psychology*, *6*, 1122. https://doi.org/10.3389/fpsyg.2015.01122

Senju, A., Kikuchi, Y., Akechi, H., Hasegawa, T., Tojo, Y., & Osanai, H. (2009). Brief report: Does eye contact induce contagious yawning in children with autism spectrum disorder? *Journal of Autism and Developmental Disorders*, *39*(11), 1598–1602. https://doi.org/10.1007/s10803-009-0785-5

Senju, A., Maeda, M., Kikuchi, Y., Hasegawa, T., Tojo, Y., Osanai, H., & Osani, H. (2007). Absence of contagious yawning in children with autism spectrum disorder. *Biology Letters*, *3*(6), 706–708. https://doi.org/10.1098/rsbl.2007.0337

Sepulveda, W., & Mangiamarchi, M. (1995). Fetal yawning. *Ultrasound in Obstetrics and Gynecology*, *5*(1), 57–59. https://doi.org/10.1046/j.1469-0705.1995.05010057.x

Sforza, C., Mapelli, A., Galante, D., Moriconi, S., Ibba, T. M., Ferraro, L., & Ferrario, V. F. (2010). The effect of age and sex on facial mimicry: A three-dimensional study in healthy adults. *International Journal of Oral and Maxillofacial Surgery*, *39*(10), 990–999. https://doi.org/10.1016/j.ijom.2010.05.011

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired

interoception. *Cortex*, *81*, 215–220. https://doi.org/10.1016/j.cortex.2016.03.021

Sheppard, E., Pillai, D., Wong, G. T. L., Ropar, D., & Mitchell, P. (2016). How easy is it to read the minds of people with autism spectrum disorder? *Journal of Autism and Developmental Disorders*, *46*(4), 1247–1254. https://doi.org/10.1007/s10803-015-2662-8

Sherman, G. D., & Haidt, J. (2011). Cuteness and disgust: The humanizing and dehumanizing effects of emotion. *Emotion Review*, *3*(3), 245–251. https://doi.org/10.1177/1754073911402396

Shmaya, Y., Eilat-Adar, S., Leitner, Y., Reif, S., & Gabis, L. V. (2017). Meal time behavior difficulties but not nutritional deficiencies correlate with sensory processing in children with autism spectrum disorder. *Research in Developmental Disabilities*, *66*, 27–33. https://doi.org/10.1016/j.ridd.2017.05.004

Smith, A. (2009). The empathy imbalance hypothesis of autism: A theoretical approach to cognitive and emotional empathy in autistic development. *The Psychological Record*, *59*, 489–510. https://doi.org/10.1007/BF033

Som, C. V. (2005). Nothing seems to have changed, nothing seems to be changing and perhaps nothing will change in the NHS: Doctors' response to clinical governance. *International Journal of Public Sector Management*, *18*(5), 463–477. https://doi.org/10.1108/09513550510608903

Song, Y., Nie, T., Shi, W., Zhao, X., & Yang, Y. (2019). Empathy impairment in individuals with autism spectrum conditions from a multidimensional perspective: A meta-analysis. *Frontiers in Psychology*, *10*(1902). https://doi.org/10.3389/fpsyg.2019.01902

Sonnby–Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology*, *43*(5), 433–443. https://doi.org/10.1111/1467-9450.00312

Sorensen, T. (2017). The role of empathy, clinical trials, and eye gaze in contagious yawning and itching. *Senior Theses and Projects*. Retrieved from https://digitalrepository.trincoll.edu/theses/643

Sovner, R., & Hurley, A. D. (1986). Four factors affecting the diagnosis of psychiatric disorders in

mentally retarded persons. *Psychiatric Aspects of Mental Retardation Reviews*, *5*(9), 45–49. Retrieved from https://psycnet.apa.org/record/1987-07641-001

Sparrow, S. S., Balla, D., & Cicchetti, D. V. (1984). *Vineland Adaptive Behaviour Scale* (Survey edn). Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Cicchetti, D. V, & Balla, D. A. (2005). *Vineland adaptive behavior scales: Second edition (Vineland II), survey interview form/caregiver rating form.* (2nd ed.). Livonia, MN: Pearson Assessment. Retrieved from https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-79948-3_1602

Stewart, M. E., & Austin, E. J. (2009). The structure of the autism-spectrum quotient (AQ): Evidence from a student sample in Scotland. *Personality and Individual Differences*, *47*(3), 224–228. https://doi.org/10.1016/j.paid.2009.03.004

Stotland, E. (1969). Exploratory investigations of empathy. *Advances in Experimental Social Psychology*, *4*, 271–314. https://doi.org/10.1016/S0065-2601(08)60080-5

Stotz-Ingenlath, G. (2000). Scientific contribution epistemological aspects of Eugen Bleuler's conception of schizophrenia in 1911. *Medicine, Health Care and Philosophy*, *3*, 153–159.

Strayer, J. (1990). Affective and cognitive processes in empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (p. 406). Cambridge: Cambridge University Press.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. https://doi.org/10.1037/a0016973

Sutherland, R. J., & Isherwood, T. (2016). The evidence for easy-read for people with intellectual disabilities: A systematic literature review. *Journal of Policy and Practice in Intellectual Disabilities*, *13*(4), 297–310. https://doi.org/10.1111/jppi.12201

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science*, *16*(6), 311–315.

Tan, C. D. (2018). “I’m a normal autistic person, not an abnormal neurotypical”: Autism spectrum disorder diagnosis as biographical illumination. *Social Science and Medicine*, *197*, 161–167. https://doi.org/10.1016/j.socscimed.2017.12.008

Teal, M. B., & Wiebe, M. J. (1986). A validity analysis of selected instruments used to assess autism. *Journal of Autism and Developmental Disorders*, *16*(4), 485–494.

The British Psychological Society (BPS). (2003). *Issues in Screening*. Leicester.

The British Psychological Society (BPS). (2015). *Guidance on the assessment and diagnosis of intellectual disabilities in adulthood*.

The National Autistic Society. (2016). The autism employment gap: Too much information. The National Autistic Society. Retrieved from http://www.autism.org.uk/~/media/nas/get-involved/tmi/tmi employment report 24pp web.ashx?la=en-gb

Timmons, J. C., Hall, A. C., Bose, J., Wolfe, A., & Winsor, J. (2011). Choosing employment: Factors that impact employment decisions for individuals with intellectual disability. *Intellectual and Developmental Disabilities*, *49*(4), 285–299. https://doi.org/10.1352/1934-9556-49.4.285

Tonnsen, B. L., Boan, A. D., Bradley, C. C., Charles, J., Cohen, A., & Carpenter, L. A. (2016). Prevalence of autism spectrum disorders among children with intellectual disability. *American Journal on Intellectual and Developmental Disabilities*, *121*(6), 487–500. https://doi.org/10.1352/1944-7558-121.6.487

Topper, S., Ober, C., & Das, S. (2011). Exome sequencing and the genetics of intellectual disability. *Clinical Genetics*, *80*(2), 117–126. https://doi.org/10.1111/j.1399-0004.2011.01720.x

Townsley, R., Rodgers, J., & Folkes, L. (2003). Getting informed: Researching the production of accessible information for people with learning disabilities. *Journal of Integrated Care*, *11*(3), 39–43. https://doi.org/10.1108/14769018200300034

Tractenberg, R. E., Yumoto, F., Jin, S., & Morris, J. C. (2010). Sample size requirements for training to a kappa agreement criterion on clinical dementia ratings. *Alzheimer Disease and Associated*

Disorders, *24*(3), 264–268. https://doi.org/10.1097/WAD.0b013e3181d489c6

Trevisan, D. A., Bowering, M., & Birmingham, E. (2016). Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions. *Molecular Autism*, *7*, 46. https://doi.org/10.1186/s13229-016-0108-6

UK Copyright Service. (2009). P-27: Using the work of others. Retrieved July 13, 2017, from https://copyrightservice.co.uk/copyright/p27_work_of_others

Underwood, L., McCarthy, J., Chaplin, E., & Bertelli, M. O. (2015). Assessment and diagnosis of psychiatric disorder in adults with autism spectrum disorder. *Advances in Mental Health and Intellectual Disabilities*, *9*(5), 222–229. https://doi.org/10.1108/AMHID-05-2015-0025

Usui, S., Senju, A., Kikuchi, Y., Akechi, H., Tojo, Y., Osanai, H., & Hasegawa, T. (2013). Presence of contagious yawning in children with autism spectrum disorder. *Autism Research and Treatment*, *2013*, 1–8. https://doi.org/10.1155/2013/971686

Van Baaren, R. B., Holland, R. W., Kawakami, K., & Van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, *15*(1), 71–74. Retrieved from http://www.despair.com

Vanvuchelen, M., Roeyers, H., & de Weerdt, W. (2011). Do imitation problems reflect a core characteristic in autism? Evidence from a literature review. *Research in Autism Spectrum Disorders*, *5*(1), 89–95. https://doi.org/10.1016/J.RASD.2010.07.010

Waight, M., & Oldreive, W. (2020). Investigating accessible information formats with people who have learning disabilities. *Learning Disability Practice*, *23*(2), 23–30. https://doi.org/10.7748/ldp.2020.e2031

Walusinski, O. (2006). Yawning: Unsuspected avenue for a better understanding of arousal and interoception. *Medical Hypotheses*, *67*(1), 6–14. https://doi.org/10.1016/J.MEHY.2006.01.020

Ward, J., Burckhardt, V., & Holle, H. (2013). Contagious scratching: Shared feelings but not shared body locations. *Frontiers in Human Neuroscience*, *7*(122), 1–2. https://doi.org/10.3389/fnhum.2013.00122

Warner, G., Cooper, H., & Cusack, J. (2019). *Autism research funding landscape A review of the in the United Kingdom Contents*. Retrieved from https://www.autistica.org.uk/downloads/files/Autistica-Scoping-Report.pdf

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition*. San Antonio, TX: Pearson.

Westbury, H. R., & Neumann, D. L. (2008). Empathy-related responses to moving film stimuli depicting human and non-human animal targets in negative circumstances. *Biological Psychology*, *78*(1), 66–74. https://doi.org/10.1016/J.BIOPSYCHO.2007.12.009

Wheelwright, S., Auyeung, B., Allison, C., & Baron-Cohen, S. (2010). Defining the broader, medium and narrow autism phenotype among parents using the autism spectrum quotient (AQ). *Molecular Autism*, *1*, 1–9. https://doi.org/10.1186/2040-2392-1-10

Wigham, S, Robertson, J., Emerson, E., Hatton, C., Elliott, J., McIntosh, B., … Joyce, T. (2008). Reported goal setting and benefits of person centred planning for people with intellectual disabilities. *Journal of Intellectual Disaabilities*, *12*(2), 143–152. https://doi.org/10.1177/1744629508090994

Wigham, Sarah, Rodgers, J., Berney, T., Le Couteur, A., Ingham, B., & Parr, J. R. (2019). Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism*, *23*(2), 287–305. https://doi.org/10.1177/1362361317748245

Wing, L. (1981). Asperger's syndrome: A clinical account. *Psychological Medicine*, *11*(1), 115–129. https://doi.org/10.1017/S0033291700053332

Witecy, B., & Penke, M. (2017). Language comprehension in children, adolescents, and adults with down syndrome. *Research in Developmental Disabilities*, *62*, 184–196. https://doi.org/10.1016/J.RIDD.2017.01.014

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart1. *Journal of Applied Behavior Analysis*, *11*(2), 203–214. https://doi.org/10.1901/jaba.1978.11-203

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Woodgate, R. L., Ateah, C., & Secco, L. (2008). Living in a world of our own: The experience of parents who have a child with autism. *Qualitative Health Research*, *18*(8), 1075–1083. https://doi.org/10.1177/1049732308320112

Woolley, H., Stein, A., Forrest, G. C., & Baum, J. D. (1989). Imparting the diagnosis of life threatening illness in children. *British Medical Journal*, *298*, 1623–1626. https://doi.org/10.1136/bmj.298.6688.1623

World Health Organisation [WHO]. (1990). *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organisation.

World Health Organisation [WHO]. (2018). *No The ICD-11 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organisation.

Yoshimura, S., Sato, W., Uono, S., & Toichi, M. (2015). Impaired overt facial mimicry in response to dynamic facial expressions in high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *45*(5), 1318–1328. https://doi.org/10.1007/s10803-014-2291-7

Zhang, Y., Zhang, L., Neoh, S. C., Mistry, K., & Hossain, M. A. (2015). Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles. *Expert Systems with Applications*, *42*(22), 8678–8697. https://doi.org/10.1016/j.eswa.2015.07.022

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS ® implementations. In *Proceeings Nesug*. Retrieved from https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf

Ziviani, J., Lennox, N., Allison, H., Lyons, M., & Del Mar, C. (2004). Meeting in the middle: Improving communication in primary health care consultations with people with an intellectual disability. *Journal of Intellectual and Developmental Disability*, *29*(3), 211–225.

https://doi.org/10.1080/13668250412331285163