The Development and Preliminary Validation of a Teacher Assessment of Barriers to Learning for Pupils on the Autism Spectrum in Special Schools

Melanie Howell

Tizard Centre

School of Social Policy, Sociology and Social Research

University of Kent

A thesis submitted for the degree of

Doctor of Philosophy

Intellectual and Developmental Disabilities

April, 2021

Words 89,079

Acknowledgements

There are so many people to thank for helping me to get this far and, although I can't thank everybody, there are a number of important people who deserve to be mentioned.

Firstly, thanks to my supervisors, Professor Peter Langdon and Dr Jill Bradshaw, for giving me the opportunity to conduct this research. Thank you for your support and guidance, for pushing me to explore new directions, for believing in me when my confidence lacked and for your endless patience.

A huge thank you must go to the participants of this research - the teachers who gave up their valuable time, the parents who allowed their children to participate and the children and young people who were the first to be assessed using the ABLE-Autism. I am hugely grateful to all of you and this research would not have been possible without you.

Thanks must also go to my fellow PhD students, particularly the few who have consistently been there to allow me to 'share the trauma' along the way. Reminding me that it's all part of the process has been so valuable in getting me through those difficult patches.

To those 21 incredible boys who I have had the honour and privilege to teach since I qualified as a teacher nine years ago. I'm sure you have taught me far more than I could ever teach you and I can't thank you or your families enough.

There is so much I owe to my family. Thanks to my Mum and Dad who have helped me reach this point in my education in a variety of ways. I certainly wouldn't be submitting my PhD without you and you can probably claim some credit for the genes! To my almost husband, Vern (who would be my actual husband by now if it wasn't for the global pandemic). Thank you for all your help, for allowing me to be a financial black hole and for always seeing the best in me. To my wonderful Grandad, who isn't here to read my PhD but asked so much about my research in the early stages and gave such wonderful advice without even knowing it. Although you would have excelled at something like this, you spent your whole life making sure that we would be able to have the opportunities that you didn't have. I was so lucky to have you and I hope I've made you proud. Finally, to my Chris. None of this would have happened without you. You have made me the person I am in so many ways and taught me more than anyone ever could. I will always be the luckiest person in the world to be able to call myself your sister.

Abstract

Individuals on the autism spectrum with coexisting intellectual disabilities are more likely to have communication difficulties, display certain aspects of restricted and repetitive sensory and motor behaviours, show adaptive behaviour impairments and display behaviour that challenges than those with either condition alone. The interactions between the two conditions result in educational needs and barriers to learning specific to this group which are not thoroughly assessed using current assessment tools. Instead, autistic pupils who have intellectual disabilities and attend special schools are routinely assessed using generic assessments appropriate for pupils with a wide range of special needs and developmental disabilities.

This thesis sought to consider the most appropriate ways to identify and monitor barriers to learning for pupils on the autism spectrum with intellectual disabilities in special schools through effective and robust assessment. Four related studies were conducted to review the available assessments appropriate for this purpose, leading to the development of a new, autism-specific, assessment to be used by teachers in special schools: The Assessment of Barriers to Learning in Education – Autism (ABLE-Autism).

Stage 1 of the research consisted of a systematic review which identified and evaluated existing assessments which could be used by teachers in special schools to measure outcomes and progress in adaptive behaviour, behaviour that challenges and autism-related behaviour for pupils on the autism spectrum. The review identified a lack of robust assessments appropriate for these purposes and justified the development of an autism-specific assessment which could be used by teachers in special schools to assess barriers to learning for these pupils.

Stage 2 included two assessment development studies. Focus groups were conducted with special needs teachers to refine the definition of the 'barriers to learning' assessment construct, consider the relevant assessment content and inform the format and features of the ABLE-Autism. A two-round modified Delphi study was then conducted in which special needs teachers endorsed 70 items for inclusion in the ABLE-Autism. The Delphi study provided initial evidence for face and content validity of the new assessment.

Preliminary validity and reliability evaluations of the ABLE-Autism were then undertaken in Stage 3. Test-retest reliability was found to be excellent, the subscales and assessment as a whole showed a high degree of internal consistency reliability and the ABLE-Autism was strongly correlated with the Teacher Autism Progress Scale. Teacher feedback was extremely positive, reflecting the fact that teachers were consulted during all stages of the assessment development process.

This research identified the need for autism-specific assessments to assess autistic pupils with coexisting intellectual disabilities in special schools. A new, robust assessment, the ABLE-Autism, was then developed for teachers in special schools to identify and monitor barriers to learning for these pupils. This research provides a practical example of how teachers can be involved in assessment development processes and also highlights that the inclusion of special needs teachers in assessment development is likely to contribute to the validity and reliability of new assessments.

Publications

Howell, M., Bradshaw, J., & Langdon, P. E. (2021). A Systematic review of behaviour-related outcome assessments for children on the autism spectrum with intellectual disabilities in education settings. *Review Journal of Autism and Developmental Disorders, 8,* 67-91. https://doi.org/10.1007/s40489-020-00205-y (Appendix 1.1)

Howell, M., Bradshaw, J., & Langdon, P. E. (2020a). 'There isn't a checklist in the world that's got that on it': Special needs teachers' opinions on the assessment and teaching priorities of pupils on the autism spectrum. *Journal of Intellectual Disabilities.* Advance online publication. https://doi.org/10.1177/1744629520972901 (Appendix 1.2)

Howell, M., Bradshaw, J., & Langdon, P. E. (2020b). Asking the teachers: A Delphi study on the selection of skills and behaviours for an assessment of barriers to learning for pupils on the autism spectrum with intellectual disabilities. *British Journal of Learning Disabilities*. Advance online publication. https://doi.org/10.1111/bld.12350 (Appendix 1.3)

Howell, M., Bailey, T., Bradshaw, J., & Langdon, P. E. (submitted). The preliminary validity and reliability evaluation of the Assessment of Barriers to Learning in Education – Autism. (Appendix 1.4)

Terminology and Definitions

Autism

The International Classification of Diseases – 10th Edition (ICD-10) refers to the conditions of "childhood autism" and "atypical autism" categorised under the umbrella of "pervasive developmental disorders" (World Health Organization [WHO], 2015, F84). Within the proposed 11th Edition (ICD-11), autism is referred to as "Autism Spectrum Disorder" with or without functional language impairment or intellectual disabilities (WHO, 2020). This is in line with the official diagnosis of "Autism Spectrum Disorder" in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V), often abbreviated to ASD (American Psychiatric Association [APA], 2013, p.31). It is recognised that there has been a social movement over recent years to recognise conditions such as autism as differences rather than disabilities and that 'condition' is preferred over 'disorder' when describing autism. In this research the term 'disorder' will be avoided where possible, apart from when directly quoted or explicitly referred to in another research paper.

There has also been much discussion in recent years on the preference of the language used to describe individuals who have a diagnosis of autism. An online study by Kenny et al. (2016) concluded that there is no universally accepted way to describe autism. The study suggested that, of the 502 participating adults who categorised themselves as autistic, the majority endorsed the terms 'autistic' or 'on the autism spectrum'. However, the data also showed that a percentage of autistic participants (as well as family members and professionals) endorsed and showed preference for other terms such as 'has autism'. No proffered terminology received a mode score of below the category of 'neither like or dislike' from the group of autistic participants. Notably, in the context of this thesis, individuals on the autism spectrum with coexisting intellectual disabilities are unlikely to have been able to access the online surveys used in the study and therefore would not have been included in such research or given the opportunity to offer their views. In line with the majority preference of those surveyed in Kenny et al. (2016), the terms 'on the autism spectrum' and 'autistic' will mainly be used in this thesis with the understanding that other terminology may be preferred.

Intellectual Disabilities

'Intellectual disability' is a term used in the DSM-V (APA, 2013). Within the ICD-10, the outdated and often pejorative "Mental Retardation" is used (WHO, 2015), although this is likely to change to "Disorders of Intellectual Development" in the ICD-11, which will officially come into effect in 2022 (WHO, 2020). In education settings in the United Kingdom, intellectual disabilities are more usually referred to as 'learning disabilities' and this appears to be the preferred term in these contexts (Cluley, 2018). However, for clarity, this research will use the term 'intellectual disabilities' to ensure that there is not confusion with specific learning disabilities. Some consultations with self-advocacy groups in England have found that people with intellectual disabilities do not like being referred to by acronyms such as ID or LD (Journal of Applied Research in Intellectual Disabilities, 2019). The use of acronyms will, therefore, be avoided in this work where possible.

Behaviour that Challenges

The concept of behaviour which is of a challenging nature has also evolved over recent years, particularly in relation to individuals with intellectual and developmental disabilities. The National Institute for Health and Care Excellence guidelines (2018) used the terminology 'behaviour that challenges' in response to the term 'challenging behaviour' being misused as a diagnosis or a description of an individual (e.g. "he has challenging behaviour"). In line with the NICE guidelines, the term 'behaviour that challenges' will be used as far as possible in this research. Behaviour that challenges has been referred to by terminology including problem behaviour, maladaptive behaviour, disordered behaviour or dysfunctional behaviour and these terms will be avoided where possible. Occasionally this terminology may be utilised in reference to the content, names or descriptions of other papers or assessments.

Accepted terminology is continually updated and revised and, although care has been taken to include acceptable terminology, it is understood that terminology used in this thesis is also likely to be considered outdated in future.

Definitions

In the course of this research, individuals on the autism spectrum with coexisting intellectual disabilities are the group of interest. When considering these individuals in a special school context, the term ‘autism’ or ‘autistic’ may be used and the coexisting intellectual disabilities are implied. If autistic individuals *without* intellectual disabilities are being referred to, this will be explicitly stated.

Similarly, this research is concerned with special education. Therefore, depending on the context, special needs teachers or special schools may be referred to as ‘teachers’ or ‘schools’ and the special education will be implied through the context. Where mainstream teachers or mainstream schools are being referred to, this will be explicitly stated.

Contents

List of Tables

List of Figures

List of Appendices

Introduction

Individuals on the autism spectrum with coexisting intellectual disabilities often have specific needs and difficulties due to the complex interaction between the two conditions. These may include communication difficulties, restricted and repetitive behaviour and interests, sensory needs and behaviour that challenges. Autistic children and young adults with the most severe and complex needs are educated in special schools. As a special needs teacher of children on the autism spectrum with severe learning disabilities, I noticed that the needs and abilities of this group of pupils often seemed disparate from the curriculum. In addition, the assessments used were usually developed for all pupils with a broad range of special needs and did not take account of some of the distinctive characteristics and needs of autistic pupils. The lack of autism-specific assessments available for use in special schools led me to question whether some of the key needs for this group of pupils, including the important progress in their learning and skills, could be better identified and recorded.

This thesis sought to consider the most appropriate ways to identify and monitor barriers to learning for pupils on the autism spectrum with intellectual disabilities in special schools through effective and robust assessment. A review of the literature is laid out in the first two chapters and four studies are then outlined which were conducted to identify and evaluate available assessments and to then develop a new, autism-specific, special school assessment: The Assessment of Barriers to Learning in Education – Autism (ABLE-Autism).

In Chapter 1, the two conditions of autism and intellectual disabilities are described. The key needs of individuals with these coexisting conditions are then considered and the differences between this group and those with either intellectual disabilities or autism alone are examined. The unique needs of this group and the potential impact upon their learning are explored. Theories of autism are outlined and the ways they may account for the needs of this population are described.

Chapter 2 has a focus on assessment. The current context of assessment in special schools is laid out and the types, methods and purposes of assessment are then summarised. The difficulties and challenges of assessing autistic pupils with intellectual disabilities are explored.

Assessment development and evaluation processes are then considered and relevant psychometric theory is described.

Chapter 3 contains a detailed description of the PhD research. The research methods are described and justified and the epistemological position of the research is outlined. Some limitations of the different research methods are considered and the ways that these limitations have been accounted for and mitigated are summarised.

To identify and evaluate the available assessments which were appropriate for assessing these pupils in special schools, a systematic review was conducted and this is described in Chapter 4. The findings of the systematic review justified the need for the development of a robust, autism-specific teacher assessment which could be used by teachers in special schools.

The assessment development process is outlined in Chapters 5 and 6. The focus groups conducted with special needs teachers are described in Chapter 5. The teacher participants identified important progress and barriers to learning for autistic pupils, along with key areas to assess and assessment features they found helpful and unhelpful. Following the focus groups, a list of skills and behaviours was created and these were presented to special needs teachers in a two-round Delphi exercise which is described in Chapter 6. The Delphi study was used to determine the items for inclusion in the assessment.

Chapter 7 contains a description of the ABLE-Autism. The construct is further defined and the purpose of the assessment is detailed. The chapter includes a description of the structure and format of the ABLE-Autism, the assessment method, specific features of the assessment and the assessment scoring.

The evaluation of the ABLE-Autism is described in Chapter 8. In this study, various psychometric properties of the assessment were evaluated. Special needs teachers who used the assessment gave feedback on aspects they found useful and the parts they thought could be improved.

Finally, the overall discussion of the research is incorporated in Chapter 9. The research and practice implications of each individual study are considered and the strengths and

weaknesses of the research are discussed. Directions of future research are suggested and recommendations made.

Chapter 1. Autism and Intellectual Disabilities – Description, Theories and Education

Within this initial chapter, autism will be introduced and some of the key characteristics of the condition will be outlined. There will be a focus upon children on the autism spectrum with coexisting intellectual disabilities, as the interaction between both autism and intellectual disabilities within educational settings leads to specific barriers to learning which may be different to barriers for those with either intellectual disabilities or autism alone. Key theories of autism will be detailed and the ways that they may account for some of the needs and difficulties of autistic individuals with coexisting intellectual disabilities within education settings will be discussed. This chapter will provide the basis for recognising autistic individuals with coexisting intellectual disabilities as a distinct group in order to ensure that educational assessment is relevant to their strengths, needs and challenges.

1.1 Autism

1.1.1 Autism History and Background

Autism, first described by Leo Kanner in the mid twentieth century (Kanner, 1943), was once thought to be a childhood disorder but is now understood to be a lifelong developmental condition (Bancroft et al., 2012). Autism can be diagnosed as early as two years old and although the mean age for diagnosis in the United Kingdom is between five to seven years, increasing diagnoses in adulthood are likely to skew this average (Brett et al., 2016). Autistic characteristics are often first noticed when language delay becomes evident (Mandell et al., 2005) and the condition is diagnosed through behavioural observation by a multi-disciplinary diagnostic assessment team (Dover and Le Couteur, 2007). Causes of autism are still unclear, however a mixture of genetic and environmental factors is thought to contribute (Amaral, 2017; Masi et al., 2017). It is difficult to know the true prevalence of autism due to differences in methodology, categorisation and data collection in prevalence studies (Chiarotti and Venerosi, 2020). Systematic reviews and meta-analyses have estimated the global prevalence of autism to be slightly less than 1%, however this may be higher in high-income countries and regions (Baxter et al., 2015). Social and political awareness of the condition has been increasing over a number

of years and this is particularly evident with the Autism Act (2009) being the first disability-specific legislation passed in England (National Autistic Society, 2013; Marshall-Tate, 2019).

1.1.2 Autism Description and Heterogeneity

Happé and Frith (1996) described autism as a "developmental disorder with a biological basis and a behavioural definition" (p.1377). Lorna Wing and Judith Gould defined autism in the 1970s as a triad of impairments (Wing and Gould, 1979). The triad comprised of deficits in social interaction, social communication and social imagination, the latter of which is now often identified through restricted and repetitive behaviours and interests (RRBs; Wing and Potter, 2002). The 10th Edition of the International Classification of Diseases (ICD-10; World Health Organization [WHO], 2015) defines "childhood autism" as:

> "(a) the presence of abnormal or impaired development that is manifest before the age of three years, and
> (b) the characteristic type of abnormal functioning in all the three areas of psychopathology: reciprocal social interaction, communication, and restricted, stereotyped, repetitive behaviour.
> In addition to these specific diagnostic features, a range of other nonspecific problems are common, such as phobias, sleeping and eating disturbances, temper tantrums, and (self-directed) aggression" (WHO, 2015, F84).

In addition to the aspects of autism specified in the ICD-10, the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) also noted that autism spectrum disorder is characterised by "persistent deficits in social communication and social interaction across multiple contexts, including deficits in social reciprocity, nonverbal communicative behaviors used for social interaction, and skills in developing, maintaining, and understanding relationships" (American Psychiatric Association [APA], 2013, p.31). The DSM-V consolidated the conditions of Asperger's Syndrome, pervasive developmental disorder and autistic disorder (both typical and atypical) into one category of autistic spectrum disorder.[1] The DSM-V allowed for specification of autism "with or without accompanying intellectual impairment" and "with or without

[1] For individuals not displaying restricted or repetitive interests, a condition of Social (Pragmatic) Communication Disorder has been created and it is likely that those with previous diagnosis of pervasive developmental disorders may fall within the criteria of this new disorder (Masi et al., 2017).

accompanying language impairment" (APA, 2013, p.51). The ICD-11 will come into effect in 2022 and is likely to align with the DSM-V, proposing 'subcategories' of autism depending on intellectual ability and functional language (WHO, 2020).

As autism is a spectrum condition, functioning may vary greatly from individual to individual (Gupta and Singhal, 2009) and this may be even more evident in the broader criteria which has resulted from the merging of the different autism diagnoses (Wing et al., 2011). Masi et al. (2017) described heterogeneity as a "hallmark" of autism (p.187). The heterogeneity of autism may be displayed in the variation of manifestations of autistic characteristics and behaviours, cognitive abilities and levels of communication, all of which may be affected by factors such as age, Intelligence Quotient (IQ), genetics and coexisting conditions (Georgiades et al., 2013). Georgiades et al. (2013) acknowledged that even though the presentation of autism can be so varied, individuals can still be classified under the predominant category of autistic spectrum condition. They, therefore, recognised autism as "perhaps a classic example of a heterogeneous disorder in which dissimilar parts are somehow connected" (Georgiades et al., 2013, p.124). As a spectrum condition, autism affects each individual differently and impacts skills, abilities and behaviours in many different ways. As mentioned above, whilst autism has been collapsed into a single unitary construct in the proposed criteria in the ICD-11 as in the DSM-V, differences in intellectual functioning and use of language have been reflected in the creation of new diagnostic codes reflecting aspects of heterogeneity in presentation.

1.1.3 Intellectual Disabilities and Models of Disability

The proposed criteria in the ICD-11 labels intellectual disability as "Disorders of Intellectual Development" and describes these as "as a group of etiologically diverse conditions originating during the developmental period characterized by significantly below average intellectual functioning and adaptive behavior" (WHO, 2020). Risk factors for intellectual disabilities include biological, neural, social and educational factors which may occur pre-, peri- or postnatal (Schalock et al., 2019). Although the amount of support required by individuals is now given more consideration when subclassifying levels of intellectual disabilities, full scale IQ remains the primary basis for coding and categorising intellectual disabilities (Schalock, 2011). Surprisingly, although the term 'intellectual disability' has been widely used for some time, it

has only recently replaced outdated terminology in the DSM-IV and ICD-10 (Carulla et al. 2011).

There has been a change in the last few decades from considering intellectual ability or disability as absolute and unchangeable to a multidimensional construct which can be supported in order to improve personal functioning and outcomes for individuals (Schalock, 2011). This has reflected the move away from the individual or medical model of disability, which focused on an individual and their impairment, towards the social model of disability, which emphasises society's role in defining a person's disability. Although some proponents for the social model of disability believe that the concept of disability is entirely socially created, many accept that disability as a construct is complex and "cannot be reduced to either social barriers or impairment" (Gallagher et al., 2014, p.1130). There are a number of perspectives that attempt to explain intellectual disabilities including those that emphasise biological and genetic factors, intellectual or behavioural limitations, interactions between people and their environments and rights-based perspectives (Schalock et al., 2019). Schalock et al. (2018) suggested that all of these perspectives serve important purposes in explaining intellectual disabilities and therefore described a "holistic theoretical framework to understand ID and guide efforts to mitigate its impact" (p.79). They combined the biomedical, psychoeducational, sociocultural and justice perspectives on intellectual disabilities and synthesised this information to create the framework to guide intellectual disabilities practice (Schalock et al., 2018). The movement towards the social model of disability has been, and continues to be crucial for the recognition of rights along with progress of social attitudes towards individuals with disabilities.

1.1.4 Autism and Coexisting Intellectual Disabilities

Individuals on the autism spectrum with coexisting intellectual disabilities often have specific needs due to the interaction of the two conditions. As mentioned above, the DSM-V allows specification of autism with intellectual or language impairment and recognises that "individuals with autism spectrum disorder often have intellectual disability" (APA, 2013, p.31). The severity of autism is classified in the DSM-V by the level of support required (*support*, *substantial support* or *very substantial support*) (APA, 2013). Similarly, the proposed ICD-11 criteria distinguish between autism with and without disorder of intellectual development and with various levels of functional language (WHO, 2020).

The rate of intellectual disabilities among individuals on the autism spectrum is unclear. Estimates have varied from around a third of autistic individuals having coexisting intellectual disabilities (Rivard et al., 2015) to up to two thirds (Chakrabarti and Fombonne, 2005) with a range of estimates in between (Charman et al., 2011b; Idring et al., 2012). Bertrand et al. (2001) found that 49% of their sample of autistic children had an IQ < 70, however a number of children were not able to be tested and, if these children were included, the percentage of the sample with an IQ < 70 rose to 63%. This study highlighted the challenges of assessing individuals with autism and the most severe and complex intellectual disabilities. Chakrabarti and Fombonne (2005) found that 66.7% of their sample were considered to have intellectual disabilities with 38.1% in the moderate to severe range. The intellectual functioning of a number of participants could not be estimated which, again, suggests that individuals with the lowest levels of functioning may not be represented by these studies. Differences in prevalence figures across studies may also be affected by diagnostic measures used or broader inclusion criteria for samples (Rivard et al., 2015). It has been suggested that, as awareness of autism among diagnosticians and practitioners increases, the number of those diagnosed with average or above average intelligence increases (Dover and Le Couteur, 2007). More recent studies may report a lower prevalence of intellectual disabilities in the autistic population due to the broadening of the diagnostic criteria (Dunn et al., 2019).

The severity of autistic characteristics is inversely related to level of general intellectual functioning, with good outcomes in adulthood associated with a higher level of general intellectual functioning (O'Brien and Pearson, 2004). The presence of coexisting intellectual disabilities has been found to be one of the strongest indicators of outcomes for those on the autism spectrum (Rivard et al., 2015). Matson and Shoemaker (2009) posited that autistic individuals with coexisting intellectual disabilities are not only a distinct group from those with intellectual disabilities alone but also from those on the autism spectrum who do not have any coexisting conditions. This idea is consistent with findings that individuals with autism and coexisting intellectual disabilities show greater impairment in social domains than matched individuals with intellectual disabilities alone or intellectual disabilities and other comorbid conditions (O'Brien and Pearson, 2004; Matson and Shoemaker, 2009). Alongside difficulties in social and language domains, autistic individuals with coexisting intellectual disabilities also show greater impairment in adaptive behaviour (O'Brien and Pearson, 2004; Matson and

Shoemaker, 2009). In addition, the coexistence of autism and intellectual disabilities is a major risk factor for behaviour that challenges (Matson and Shoemaker, 2009). The severity of intellectual disabilities can result in those on the autism spectrum having complex needs in terms of both social care and education. For example, Arnold and Reed (2016) noted that 25% of autistic individuals are nonverbal, whereas Matsuzaki et al. (2019) suggested that 30-40% of children on the autism spectrum remain minimally verbal into adulthood. If those on the autism spectrum and coexisting intellectual disabilities have specific needs as described above, it is important that this group is recognised and considered independently in order to accurately and appropriately consider their needs (Hurley and Levitas, 2007).

In the last few decades, there has been an advance in the recognition of autistic individuals without accompanying intellectual disabilities. The identification of children on the autism spectrum in mainstream schools and the ways that they could be included in and access mainstream education, quite rightly, became a large focus of autism research (Jordan, 2001). Commentators have highlighted the importance of including the voices of those who are the focus of the research and this is becoming more common with the involvement of advocacy groups and the increase of collaborative research (Kaehne and O'Connell, 2010). However, a possible unintended effect of these improvements in inclusive research practice is that there may be less of a focus on the individuals requiring the most support or who are considered unable to participate, such as children on the autism spectrum with moderate to profound intellectual disabilities (Bal et al., 2016; Jordan, 2001). A recent review by Stedman et al. (2019) found that the proportion of intervention and outcome studies focused on autistic individuals with lower functioning and communication has been reducing over time in line with changes to the diagnostic criteria. Stedman et al. (2019) identified that this group "for whom assessment and treatment pose a particular challenge, is arguably the least well-understood" (p.1389). There are many perceived difficulties for researchers when conducting research with individuals with more severe and complex disabilities including ethical considerations and further practical complexities (Bal et al., 2016; Crook et al., 2016) and these perceived difficulties may also result in less research with a focus on these groups.

For these reasons, individuals on the autism spectrum with coexisting and intellectual disabilities who require education in special education settings will be the focus of this research.

There are differences in the level of general intellectual functioning between mild, moderate, severe and profound intellectual disabilities and these are often reported in research. However, many pupils in special schools will not have been subject to specific testing in order to determine the severity of their intellectual disabilities. Therefore, all individuals on the autism spectrum with any level of intellectual disabilities will be considered in this research if they are educated in a special school. As much consideration as possible will be given to the individuals with moderate, severe and profound intellectual disabilities as an often underrepresented group in research.

1.2 Educational Needs of Children on the Autism Spectrum with Coexisting Intellectual Disabilities

1.2.1 Pupils on the Autism Spectrum Attending Special Schools

With a push for inclusion, many autistic pupils are educated, wholly or partly, in mainstream schools. Although special educational needs (SEN) data is reported by government as a whole and, therefore, data specific to pupils on the autism spectrum is not available, recent data show over 82% of pupils with SEN attend a mainstream primary or secondary school (Department for Education [DfE], 2019a). Support may range from a pupil being an ordinary member of the class, but with specific needs and targets identified, to provision of personal one-to-one support or accessing some educational provision away from the main classroom teaching (Wearmouth and Butler, 2020). It is currently only children with the most severe and complex learning needs that require educating in a special school and this makes up only 9% of pupils with SEN (DfE, 2019a). For pupils whose needs can't be met through the SEN resources routinely provided to mainstream schools, an Education, Health and Care Plan (EHCP) can be put in place. The EHCP is a legal document outlining the needs of a pupil and the provision required. Reports for 2019 show that 29% of pupils with EHCPs have a primary additional need such as autism (DfE, 2019b). Although official figures aren't available, estimates suggest that up to 50% of pupils on the autism spectrum may be educated in special schools although other estimates suggest that the real percentage may be lower (Reed et al., 2012). With the support available to enable inclusion in mainstream schools for many pupils on the autism spectrum, it is likely that the majority of autistic pupils who are educated in special schools have some level of coexisting intellectual disabilities.

Some contend that special education is segregationist and in opposition to the social model of disability (Gallagher et al., 2014). The arguments for and against inclusion and special needs schooling are beyond the scope of this thesis. Special schools, however, allow for staff, usually with additional training, to implement individual and personalised teaching which takes account of a pupil's strengths and interests. Although it is acknowledged that practice may ultimately vary between schools, the learning environment in special schools is usually designed and adapted to support a pupil's learning needs and expectations are high with a focus on future outcomes. It is therefore argued that values in special education are very much in line with the social model of disability and include aims of supporting and enabling each pupil to overcome barriers, access learning opportunities and gain independence, regardless of their ability or disability. With the requirement of compulsory full-time education, special schools are clearly appropriate for some pupils with the most complex needs due to the specially trained staff, appropriate interventions and specialist support which is unlikely to be available to them in a mainstream setting (Reed et al., 2012).

For autistic children with intellectual disabilities who are educated in special schools, educational focus may be different to that of mainstream schools. Behavioural and social outcomes, for example, are often a large part of teaching and intervention in special schools (Reed et al., 2012) and many of the individual pupil targets link to quality of life outcomes, particularly aspects of functional independence (Schalock, 2004) and areas affecting family quality of life (Lord et al., 2020). For pupils with autism in special schools, the interaction of autism and intellectual disabilities together with the social and physical environment may result in a number of difficulties and challenges which affect access to learning, particularly when they are young. 'Barriers to learning' is a widely used term in education and refers to anything which may result in difficulties accessing learning opportunities. There are many broad aspects of the child's life which may be considered to contribute to difficulties in accessing learning including, for example, the presence of special educational needs (Cheminais, 2013), family's socio-economic status (Adelman and Taylor, 2006) and environmental barriers (Nelson and Soli, 2000). Certain barriers, such as socio-economic status, are difficult issues for a teacher to address alone and are outside of the scope of this thesis. As well as teaching skills and behaviours related to positive adult outcomes and quality of life, teachers in special schools also use and devise appropriate adjustments to the school environment in order to reduce and overcome

environmental barriers to learning. In order to create the personalised environmental adaptations and individual teaching strategies to facilitate learning and skill development, it is necessary to identify and monitor each child's difficulties and skill gaps. This research will therefore focus on behaviours and skill deficits which affect the opportunity for pupils on the autism spectrum to access and engage in learning in a special needs classroom. The skills and behaviours relevant to barriers to learning in the context of this research are ones which are able to be identified, observed and measured by special needs teachers. 'Barriers to learning' as an assessment construct will be considered in the following chapters and defined in Chapter 7. The following sections will discuss a number of challenges for many autistic individuals, the additional difficulties resulting from the complex interaction of autism and intellectual disabilities and how these may be considered barriers to learning in a special school setting.

1.2.2 Communication

Communication in its broadest sense is central to everyday life. Communication and language difficulties are at the centre of autism; the DSM-V specified that "persistent deficits" in social communication and social interaction must be present for an autism diagnosis and identified difficulties in areas including verbal and nonverbal interaction as well as understanding (APA, 2013, p.31). Severity level specified in the DSM-V ranges from atypical responses and decreased interest in social interactions at level one through to few words of intelligible speech at level three (APA, 2013).

Although it is important to note that having minimal language is not necessarily synonymous with intellectual disability (Bal et al., 2016), individuals with autism and intellectual disabilities are likely to show more limited communication and language than individuals without intellectual disabilities or with mild intellectual disabilities (O'Brien and Pearson, 2004). Autistic individuals with moderate to severe intellectual disabilities are also more likely to show abnormal language use such as echolalia, idiosyncrasy and neologisms (O'Brien and Pearson, 2004). Development of some functional language by age five is a predictor of positive outcome in adulthood (Howlin, 2005) with receptive language and social-communication impairment at two years old being a predictor of levels of independence in daily living skills as an adult (Bal et al., 2015). Research had suggested that there may be discrepancies between the expressive and receptive communication abilities of individuals with

autism, although a meta-analysis conducted in 2014 found no evidence of expressive language skills being higher than receptive skills (Kwok et al., 2015). The results of the meta-analysis did, however, identify that those with an autism diagnosis had lower receptive and expressive abilities compared with typically developing peers and it is necessary to note that the conclusions were drawn across a range of ages and cognitive abilities, not just with children with coexisting intellectual disabilities (Kwok et al., 2015).

Minimally verbal pupils are more likely to be educated in a special school setting and staff in special schools often have a focus on teaching communication throughout all subjects and across the school day, particularly in terms of teaching personalised alternative methods of communication where appropriate (Jordan, 2001). For example, a study by Abbott and Lucey in 2005 identified that 77% of 812 special schools responding to a postal questionnaire reported the use of symbols (a type of alternative and augmentative communication [AAC]) in school and this included 96% of schools for pupils with severe learning disabilities. Learning involves the ability to communicate, both receptively and expressively, and difficulties with functional communication may result in potential barriers to learning in a number of ways. Difficulties in receptive understanding will affect a child's ability to make sense of what is happening around them, understand instructions and to receive input from others. Expressive communication difficulties will affect the ability to effectively communicate with school staff and peers including the ability to express what they do know. For example, it may be harder to assess the abilities of individuals with more severe impairments in language and communication due to floor effects that may be encountered (Bal et al., 2016). As communication is an important part of everyday life, communication difficulties may have wider implications resulting in further barriers to learning. Studies have suggested that communication difficulties may be a risk factor for social anxiety (Pickard et al., 2017; Spain et al., 2018) and may serve as a function of behaviour that challenges (Bowring et al., 2017; Gore et al., 2013). Communication difficulties have a direct and indirect impact upon access to learning, particularly for pupils whose communication difficulties are as a result of the interaction between autism and intellectual disabilities.

1.2.3 Cognitive Profile

Autistic individuals with coexisting intellectual disabilities often display a unique cognitive profile. Although autism is a heterogeneous condition, since the early work by Hermelin and O'Connor (1967) unusual patterns of development and abilities have been acknowledged (Semino et al., 2018). While early research focused on the deficits and impairments associated with autism, more recent research has also acknowledged certain strengths (Pellicano et al., 2006). Specific strengths and weaknesses have been observed in the cognitive profiles and, in addition, they appear to be unique to autism and not found in individuals with non-specific intellectual disabilities (Happé and Frith, 1996; Happé, 2018). Despite intellectual disabilities or perceived low levels of ability in certain domains, individuals can display average or even above average ability in specific skills (Hermelin and O'Connor, 1967). The most known example of this type of profile is that of savant skills, defined as skill or knowledge higher than both the individual's level of general functioning and the population norm (Howlin et al., 2009). Although often thought of as extreme abilities and talents, the uneven profile can also comprise of some skills at the expected level with other abilities delayed or impaired. Even without comparing to average abilities of the population, specific skills involving rote memory, particularly of nonsense or random stimuli (Meilleur et al., 2015; Shah and Frith, 1983; Pellicano et al., 2006), and spatial visualisation (Happé, 1994) have previously been identified as peaks of ability for autistic individuals. Research has suggested that individuals on the autism spectrum often present with weaker verbal skills relative to their non-verbal skills (Rivard et al., 2015), particularly for those with coexisting intellectual disabilities (Boucher, 2017). However, Happé and Frith (1996) suggested that, contrary to expectation, communication difficulties may not fully account for weaknesses, as not all verbal subtests of the Wechsler scales show weak performance just as not all non-verbal subtests indicate strengths. The digit span test, for example, was shown to be a strength despite measuring verbal short-term memory and, in contrast, the non-verbal picture arrangement test was seen as a weakness. The “often uneven” intellectual profile has been recognised in the DSM-V (APA, 2013, p.51) and the wider interest and acknowledgement of this developmental unevenness can be seen by the different ways that this feature of autism has been described: “splinter skills, islets of abilities, special isolated skill (SIS), peaks of abilities, uneven cognitive profile, and cognitive disharmony” (Meilleur et al., 2015, p.1354).

Conversely, some recent studies are disputing the extent of this atypical developmental profile. The disparities in the findings of some studies may be due to methodological differences or variations in the sample used. Charman et al. (2011b) found evidence suggesting only a small difference between Performance IQ and Verbal IQ in contrast to previous research, although they did find some evidence of unevenness. Burack and Volkmar (1992) found similar sequences of development between both autism groups (IQ > 50 and IQ < 50) and control groups in receptive and expressive communication domains. They did, however, find evidence for uneven adaptive behaviour profiles for the individuals on the autism spectrum with IQ < 50 (Burack and Volkmar, 1992).

Although some studies such as Burack and Volkmar (1992) have explored the cognitive profiles of individuals on the autism spectrum with coexisting intellectual disabilities, the research in this area has tended to focus on those with average or slightly below average IQ. Individuals in the moderate to severe intellectual disability range are often excluded from participation in studies in order to study the "autism itself"; isolating the autism and avoiding the influence of intellectual disabilities on results (Charman et al., 2011a). If, as discussed above, those on the autism spectrum with intellectual disabilities are to be considered a distinct group with specific needs, then it is important for research to use broader samples which include such individuals in order to consider whether profiles are consistent with those with average ability. Notwithstanding the lack of research and some variation in the research findings, what appears to be the case is that a pattern of strengths and weaknesses in autistic individuals can (at a minimum anecdotally and at best evidenced through research) be determined and identified by those working with these individuals and by autistic individuals themselves (Milton, 2012). It may also be the case that for autistic individuals with coexisting intellectual disabilities, the spiky profile is more evident or pronounced. As Jordan (2001) acknowledged:

> "Children with autism and SLD [severe learning disabilities] are likely to show a developmental pattern which is both deviant and delayed; it is the deviance due to the autism that will have the greatest impact on teaching and learning but neither aspect can be ignored" (Jordan, 2001, p.11).

The splinter skills or uneven profile may affect pupils' access to learning in school in a number of ways. Pupils' abilities may be over- or underestimated, resulting in work or teacher

expectations being pitched at the wrong level. Certain prerequisite skills, for example joint attention or receptive and expressive communication, may be less developed or they may be rote learnt rather than the individuals developing a deep understanding of concepts (Koegel et al., 2012). This could result in difficulties understanding more complex concepts or applying learnt skills to other situations. As considered above, impairments in communication skills may result in difficulties understanding instructions as well as potentially leading to behaviour that challenges as a consequence of not being able to communicate wants, needs and feelings (Chiang, 2008). In addition, the linear structure of the education system where pupils complete and master certain skills or knowledge which is assessed before moving on to other learning may not suit pupils such as those on the autism spectrum whose learning is often non-linear due to uneven cognitive profiles. The uneven cognitive profile is an important factor when considering the potential barriers to learning that may be displayed by this specific group of pupils if certain skills have not been achieved that are prerequisites for other skills.

1.2.4 Restricted and Repetitive Behaviours and Interests

The atypical nature of restricted and repetitive behaviours and interests (RRBs) in autism has been recognised from the very first descriptions of autism by Kanner (1943) and now represent deficits in social imagination from Wing and Gould's triad of impairments (Wing and Potter, 2002). To receive a diagnosis of autism, there must be evidence of restricted and repetitive behaviour or interests, either current or in the individual's history (APA, 2013). It is necessary to recognise that RRBs are not unique to autism, however research does suggest that they appear to occur at higher levels or frequencies for autistic individuals than for those with other conditions (Berry et al., 2018; Leekam et al., 2011) and often are more extreme (Frith, 2003). In the context of autism, 'restricted' refers to inflexibility, intolerance of change and narrow focus whilst 'repetitive' pertains to rituals, routines and stereotypies in speech, motor movements and behaviour (Leekam et al., 2011). This set of behaviours are referred to in various diagnostic assessment tools, however more research is needed into consistency across measures (Szatmari et al., 2006).

The DSM-V requires two of the following four criteria to be met in order to this diagnostic criterion to be satisfied:

> "1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).
> 2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).
> 3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).
> 4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement)" (APA, 2013, p.50).

The effect of RRBs on an individual's level of functioning also affects the classification of the autism severity level set out in the DSM-V (APA, 2013).

Although many commentators recognise that there may be considerable overlap between categories, Szatmari et al. (2006) reflected on previous research and suggested that the area of RRBs may be broken down further into two categories: insistence on sameness, considered a higher order RRB, and repetitive sensory and motor behaviours which are considered lower order RRBs. They also found that higher scores on the insistence on sameness domain, which encompasses inflexibility and rigidity in behaviour, were more likely for autistic individuals without coexisting intellectual disabilities. Conversely, those with coexisting intellectual disabilities scored more highly for the repetitive sensory and motor behaviours, suggesting that the severity of this domain may be reflective of developmental level (Szatmari et al., 2006). Leekam et al. (2011) also recognised that the literature in this area indicated that stereotypies and preoccupation with parts of objects are more likely in young children on the autism spectrum or those with coexisting intellectual disabilities. As the diagnostic criteria recognises, this feature of autism may impact upon an individual's functioning and this impact is likely to increase with the presence of coexisting intellectual disabilities.

Although RRBs can often serve a purpose for individuals on the autism spectrum such as reducing anxiety, when they occur at particularly high frequencies or in certain forms, they can not only have considerable impact upon an individual's abilities to access learning opportunities (Berry et al., 2018; Sethi et al., 2019), but may even be harmful to the individual or others (O'Brien and Pearson, 2004). The suggested link between developmental level and the presence of repetitive sensory and motor behaviours means that these types of RRBs are more likely to be displayed by pupils in special schools who are working at lower developmental levels. Rigidity in thought or behaviour may also impact opportunities to learn. Unexpected changes or interference with preferred objects, for example, may result in anxiety and distress and lead to behaviour that challenges (Rodgers et al., 2012). It is necessary to identify that special interests can provide opportunities for teaching and learning by being incorporated into classroom practice, however this may be more relevant for autistic individuals without intellectual disabilities (Gunn and Delafield-Butt, 2016). The sensory aspect of RRBs may also impact learning opportunities and will be discussed further below. Many studies have shown that parents and teachers identify RRBs as some of the most difficult to manage as well as contributing to parent stress (Sethi et al., 2019; Welsh et al., 2019). Focus in special schools may be on providing and teaching safe and alternative forms of severe RRBs as opposed to reducing or preventing them as this may be harmful to individuals. A balanced approach is often taken depending on the impact of the behaviours on an individual's quality of life and that of their family. As they are central to the diagnosis of autism and may reflect aspects of developmental level, RRBs are key autism-related behaviours which may result in barriers to learning for children with autism and intellectual disabilities in special schools.

1.2.5 Adaptive Behaviour

Adaptive behaviour is defined as the practical behaviour which allows an individual to function in their daily life and includes the use of functional communication as well as independence, self-help and physical motor skills (Kanne et al., 2011). The mixture of behaviours which can be considered 'adaptive behaviour' means that the definition of the construct can be unclear (Kramer et al., 2012). For the purposes of this research, adaptive behaviour in school includes functional, applied or generalised skills including independence.

Research has suggested that lower adaptive behaviour is a possible predictor for reduced family quality of life (Emily and Grace, 2015). Difficulties in adaptive behaviour are usually related to level of intellectual disability, although research suggests that the presence of autism may affect adaptive behaviour beyond the expected relationship between adaptive behaviour and intellectual disabilities (Chatham et al. 2018). Individuals on the autism spectrum often display the most impairments in the communication and socialisation domains of adaptive behaviour (Bradshaw et al., 2019; O'Brien and Pearson, 2004). Autistic individuals with intellectual disabilities show greater impairments in these areas than matched individuals with intellectual disabilities alone but, similarly, those on the autism spectrum without intellectual disabilities show greater adaptive behaviour difficulties than neurotypical individuals (O'Brien and Pearson, 2004). It appears that the interaction between autism and intellectual disabilities, as mentioned when discussing autism with accompanying intellectual disabilities above, may affect an individual's ability to acquire adaptive behaviour skills in a greater way than the impact of just intellectual disabilities alone (Matson and Shoemaker, 2009).

Adaptive behaviour impacts upon many aspects of a child's ability to access learning. Difficulties with organisation and independence, for example, may make it difficult for an individual to follow both cognitive and physical school routines. More broadly, adaptive behaviour difficulties have been linked with lower academic and behavioural school achievement for typically developing children (De Bildt et al., 2005). De Bildt et al. (2005) found that children who attended special schools for pupils with more severe and complex needs showed lower adaptive behaviour skills than those attending schools for more mild learning disabilities. Although potentially encompassing a wide variation of behaviours, skills and abilities, adaptive behaviour difficulties can result in barriers to learning for pupils on the autism spectrum in special schools.

1.2.6 Behaviour that Challenges

Just as adaptive behaviour is usually seen to decrease with more severe intellectual disabilities, the likelihood of behaviour that challenges increases (O'Brien and Pearson, 2004). The concept of behaviour that challenges has changed over time. Its definition is linked closely with cultural and social expectations of behaviour that may affect the safety of both the

individual and other people as well as that which may limit an individual's access to facilities in the community (Emerson, 2001). Behaviour that challenges can be defined as follows:

> "Behaviour can be described as challenging when it is of such an intensity, frequency or duration as to threaten the quality of life and/or the physical safety of the individual or others and is likely to lead to responses that are restrictive, aversive or result in exclusion." (Royal College of Psychiatrists et al. 2007).

As mentioned in the above definition, the presence of behaviour that challenges is linked to lower quality of life scores (Beadle-Brown et al., 2009) and may have a large impact upon parental stress and family quality of life (Ludlow et al., 2012). A meta-analysis by McClintock et al. (2003) estimated 10-20% of individuals with intellectual disabilities display behaviour that challenges. Their results suggested that individuals on the autism spectrum with coexisting intellectual disabilities are more likely to show self-injury, destruction to property and aggression than those without autism. They also found that those with severe to profound intellectual disabilities were more likely to display self-injurious behaviour and stereotypy (McClintock et al., 2003). A recent study by Nicholls et al. (2020) found that, in special schools, behaviour that challenges was significantly associated with reduced adaptive behaviour and with a diagnosis of autism.

Along with cultural perceptions of behaviour that challenges, the definition may also be setting dependent or determined by impact (Oliver et al., 2003). In the context of the special needs classroom, behaviour that challenges can extend beyond the often cited aggression, self-injury and property destruction. In a school environment, behaviour that challenges can also include:

> "Behaviour which prevents participation in appropriate educational activities; isolates children from their peers; affects the learning and functioning of other pupils; drastically reduces opportunities for involvement in ordinary community activities; makes excessive demands on staff and resources; places the child or others in physical danger; and makes the possibilities for future placement difficult" (Male, 1996, p.310)

In this way, behaviour that challenges can create significant barriers to learning opportunities for autistic pupils with intellectual disabilities in special schools. Special needs teachers have previously reported prevention of learning or limiting of opportunities as particular concerns regarding behaviours that challenge, alongside the more expected responses which identified concerning behaviours such as injury to pupils and staff (Male, 2003). In addition, studies have reported that behaviour that challenges can also have a negative impact on staff emotional wellbeing, contributing to stress, anxiety and burnout (Amstad and Müller, 2020; Hastings and Brown, 2002). It is necessary to note that behaviours specifically relevant to autism such as stereotypic behaviours and resistance to change discussed above may also be considered behaviour that challenges in certain settings or contexts and depending upon intensity, frequency and duration. These behaviours, along with characteristics central to autism such as communication and language difficulties, may interact to increase the likelihood of these pupils displaying behaviour that may be considering challenging in the context of a school environment (Murphy et al., 2005).

1.2.7 Sensory Needs

Sensory needs and differences are highlighted by the DSM-V as one of the four criteria, of which two need to be present, for an autism diagnosis (APA, 2013). It has been estimated that over 80% of children on the autism spectrum have some difficulties with sensory processing (Case-Smith et al., 2015). The DSM-V identified that autistic individuals may be hyper- or hyporeactive to certain stimuli and sensory feedback and this may manifest as an unusual interest in environmental sensory input (APA, 2013). Hyperreactivity to sensory input may result in sensory defensiveness to stimuli such as loud sounds, bright or fluorescent lights, smells, textures and tastes. Hyporeactivity can cause a decreased registration of sensation and passivity to input such as pain or temperature. Hyporeactivity may also produce sensory seeking behaviours. Sensory seeking behaviours may be related to movement and motion, visual sensation, such as watching spinning objects, and smelling, licking, mouthing or eating inappropriate objects. Although sensory differences are not unique to autism and can be found in individuals with a range of developmental disabilities, research again suggests that individuals on the autism spectrum with coexisting intellectual disabilities display different sensory needs to those with intellectual disabilities or autism alone. Joosten and Bundy (2010) conducted an Australian study

and found that autistic children with intellectual disabilities were significantly more sensitive to sensory input and displayed more extreme sensory avoidance responses than children with intellectual disabilities without autism, although results did not reach significance in relation to hyperresponsiveness. As recognised in the DSM-V, research has also suggested that sensory processing difficulties are linked to RRBs for autistic children, particularly with stereotypic and repetitive movements (Joosten and Bundy, 2010; Suarez, 2012). Grzadzinski et al. (2020) have suggested that patterns of sensory reactivity may predict outcomes for children on the autism spectrum while Lane et al. (2010) found different patterns of sensory needs to be predictors of communication impairment and behaviour that challenges.

Intense sensitivities to sensory stimuli can result in functional difficulties, with children not accessing or engaging in environments which are loud, busy or bright, as well as potentially leading to anxiety and behaviour that challenges. Sensory seeking behaviours can affect individuals' engagement and focus, for example due to the desire for regular movement (Kern et al., 2007). In extreme cases these behaviours could have health implications, such as ingesting non-edible or harmful substances (Matson et al., 2011). Particular sensory needs may limit opportunities within a classroom or school environment, for example by inappropriate sensory seeking or sensory aversions affecting attention and focus in the ways described above (McAllister and Maguire, 2012; Vogel, 2008). In recent years, increased importance is being placed on an appropriate learning environment in schools with the impact of the environment being considered in the design on new buildings as well as the general set-up of classrooms (Department for Children, Schools and Families, 2014). It is now widely recognised within education that autistic pupils may be particularly affected by their learning environment (McAllister and Sloan, 2016) and the environments in special schools are specifically adapted to account for pupils' sensory needs. Adaptations in special schools include carefully structuring the classroom and school environment using furniture to define 'zones' or areas, workstations with minimal distraction, individualised sensory integration therapy provided throughout the school day, high pupil-staff ratios, interactive displays and the use of visual communication including symbols, Picture Exchange Communication System (PECS), objects of reference, visual timetables and individual schedules. For pupils on the autism spectrum with intellectual disabilities, sensory needs may be such that difficulties persist despite the adaptations made to the environment. As well as presenting difficulties in schools, unusual sensory needs may also

impact on family life by limiting participation for families in leisure activities and experiences as a family (Schaaf et al., 2011). Again, as with RRBs, the aim of teaching in special schools is often to support pupils rather than preventing sensory-related behaviour. Focus is often on teaching pupils to communicate their sensory requirements and to sensitively find and teach appropriate ways for pupils to have their sensory needs met safely whilst allowing the pupils to access learning and life experiences which may be restricted by unsupported sensory needs.

1.2.8 Summary

The interaction of autism and intellectual disabilities as coexisting conditions often results in specific difficulties for these individuals which may be different to the difficulties encountered by those with either intellectual disabilities or autism alone. Research has suggested that autistic individuals with coexisting intellectual disabilities are more likely to have more limited communication (O'Brien and Pearson, 2004), display certain aspects of RRBs including repetitive sensory and motor behaviours (Szatmari et al., 2006) and show adaptive behaviour impairments, particularly within communication and social domains (Chatham et al. 2018). Autistic individuals with coexisting intellectual disabilities are also more like to display behaviour that challenges, perhaps as a result of the difficulties and impairments in these other domains (Nicholls et al., 2020). Alongside these difficulties, the uneven cognitive profile seen in autism, which is often more pronounced for this group (Boucher, 2017), may result in underdeveloped skills and abilities in key areas or a misrepresentation of what the individual is capable of. The consequence of these difficulties, either individually or collectively, results in educational needs which are specific to the group of pupils who have coexisting autism and intellectual disabilities. Many of these difficulties can be further understood by contextualising them within current theoretical perspectives about autism, which will be considered within the next section. These areas of difficulty have an effect on family life and therefore are socially valid areas for schools to address, not only to improve access to classroom and school learning for pupils, but also to improve family quality of life and future quality of life outcomes.

1.3 Theories of Autism

There are a number of theories which attempt to explain the characteristics and features of autism, three of which dominate autism research; Empathising-Systemising Theory (E-S)

which revised the well-known Theory of Mind (ToM), Executive Function (EF) and Weak Central Coherence (WCC). These theories, along with some supplementary theories of autism, will be outlined and considered in light of the educational needs of those on the autism spectrum with coexisting intellectual disabilities discussed above.

1.3.1 Empathising-Systemising Theory

Baron-Cohen et al. (1985) described the ToM hypothesis as it relates to autism (sometimes referred to as 'metalising' or 'mind blindness') as an "inability to represent mental states" (p.43). They suggested that problems assigning thoughts and beliefs to others result in difficulties socialising as well as understanding and predicting the behaviour of other people. ToM difficulties for those on the autism spectrum has become a widely acknowledged theoretical framework explaining some of the social communication deficits found in those on the autism spectrum, although it is perhaps best thought of as an impairment or developmental delay rather than a strict inability (Frith, 2003). Happé (1994) suggested that, as a deficit-based hypothesis, ToM does not account for the strengths seen in the uneven cognitive profile discussed above. Baron-Cohen (2009) also identified weaknesses of the theory, including that it does not account for non-social features of autism such as the focus on fine detail and RRBs.

Baron-Cohen (2009) suggested a revision of this hypothesis to include an explanation of the strengths of autism, which he described as an average or above average ability in systemising. Systemising refers to the ability of an individual to seek, analyse or create rules. This theory, termed E-S, proposes that autistic individuals show impairments in empathising and superior skills in systematizing and the difference between these two abilities is indicative of autism. It offers an explanation as to why certain skills and abilities, such as spatial-visualisation, may be stronger or even exceptional in autistic individuals. The theory suggests that small variation in details of the environment may mean the learned 'system' breaks down and this may account for difficulty in adaptive ability and generalisation of skills (Baron-Cohen, 2009). As a theory, E-S goes further than ToM to present an account that recognises highly systemised behaviour potentially resulting in the rigidity, inflexibility and repetitiveness seen in RRBs (Grove et al., 2013). It was suggested in the description of E-S that it may only apply to autistic individuals without intellectual disabilities, however Baron-Cohen (2009) recognised that it

should also be possible to test abilities of both empathising and systemising in those with intellectual disabilities.

The Empathy Imbalance Hypothesis (EIH) extends E-S by further defining the concept of empathy. Smith (2009) suggested that cognitive empathy and emotional empathy are distinct but complementary domains. EIH proposes that autistic individuals may have difficulties with cognitive empathy, which includes predicting behaviour and attributing mental states, but heightened or oversensitive aspects of emotional empathy, which involves emotionally responding to the emotional states of others. This may result in difficulties with social communication and adaptive functioning as well as a reliance on predictable, controllable systems resulting in RRBs. Smith (2009) identified that a heightened or oversensitive emotional empathy may explain avoidance patterns and could also be related to sensory sensitivities seen in many autistic individuals.

1.3.2 Executive Function

A further deficit-based theory of autism involves executive function (EF). This was described by Ozonoff et al. (1991) as:

> "the ability to maintain an appropriate problem-solving set for attainment of a future goal; it includes behaviors such as planning, impulse control, inhibition of prepotent but irrelevant responses, set maintenance, organized search, and flexibility of thought and action" (p.1083).

Early studies on the relationship between EF and autism suggested that specific areas of EF, such as attention shifting, flexibility and planning, are underdeveloped for autistic individuals. Early studies described an "uneven performance" on a battery of neuropsychiatric tests as well as "inflexibility and perseveration" in problem solving tasks (Steel et al., 1984, p.706), whilst spatial analysis and non-verbal skills were intact or even above average (Rumsey, 1985; Rumsey and Hamburger, 1988). RRBs have also been considered by some commentators to be extreme manifestations of perseveration, exacerbated in individuals on the autism spectrum due to difficulties in problem solving and inabilities to shift attention (Lopez et al., 2005). At first glance, impairment in EF may seem to account for many behavioural and cognitive characteristics of autism, including the uneven profile and RBBs. However, research has

sometimes shown inconsistent results when considering variations in sample (for example, age and IQ) and dependent upon the specific domains of EF being tested (Hill, 2004). It is necessary to recognise that impairments in EF are also found in those with a number of other intellectual and developmental disabilities (Hughes, 1998). Although, again, studies in this area often exclude autistic participants with coexisting intellectual disabilities, Tsermentseli et al. (2018) conducted research which looked at the EF profile of autistic children with intellectual disabilities. They examined the correlation of EF with social impairment and adaptive behaviour and found similar results to studies assessing the EF profile of autistic individuals without intellectual disabilities. Findings indicated that EF correlated with adaptive functioning but not social impairment. However, it was acknowledged that comparison with other groups, such as those with intellectual disabilities alone, would be useful and that further research considering autistic individuals with intellectual disabilities is necessary (Tsermentseli et al., 2018). It may be that individuals on the autism spectrum show impairment in EF in different ways to individuals with other developmental disabilities (Hill, 2004) or perhaps that executive dysfunction interacts with other impairments, such as those described by ToM deficits above, to create an autism phenotype.

1.3.3 Weak Central Coherence

The theory which seeks to explain the uneven cognitive profile seen in autism is WCC. Central coherence is described by Happé (1999) as "the everyday tendency to process incoming information in its context – that is, pulling information together for higher-level meaning – often at the expense of memory for detail" (p.217). Shah and Frith (1983) examined the performance of autistic individuals on the embedded figures test and compared it to typically developing individuals and individuals with intellectual disabilities but without autism. Autistic individuals achieved age appropriate scores on this task even though their functioning was lower than their chronological age. They, therefore, hypothesised that individuals on the autism spectrum may find an embedded figure easily and quickly because the whole figure appeared less dominant (Shah and Frith, 1983). Frith (2003) described this "effect of context on meaning" as central coherence (p.152). WCC, then, suggests that individuals with autism have impairments in central processing and, as a result, have a bias towards detail and local information (Happé and Frith, 2006). This bias results in a failure to understand context, generalise skills and see "the big

picture" (Happé and Frith, 2006). This may explain the difficulties that those on the autism spectrum face understanding nuances of language and communication. It may also account for peaks of ability in tasks requiring detailed visual perception such as puzzles and block design tests as well as accounting for adaptive behaviour difficulties associated with generalisation skills. Many abilities associated with savants include arithmetic or calculations, music, art and memory (Howlin et al., 2009) which Happé (1999) suggested may be accounted for by the cognitive style sensitive to detail as described by WCC. It may also offer some explanation towards the disparity between procedural memory, in which sequences and details can be remembered, and declarative memory, parts of which require more contextual understanding (Semino et al., 2018).

WCC may also explain restricted behaviour seen in autism. Frith (2003) proposed that the focus on detail by those with autism may explain restricted interests and rigidity may be a product of the lack of awareness of context needed to act and respond flexibly. The focus on detail could also result in small, specific actions being more recognisable and hence more replicable. What is not explained by WCC is why repetitive behaviour occurs. Frith (2003) suggested that perhaps the inhibition of 'automatic' repetitive movements is affected by the difficulties with social communication, whereas neurotypical individuals are likely to control these types of socially undesirable repetitive behaviours due to social awareness (Frith, 2003). This may suggest WCC has links and overlaps with executive dysfunction which will be further considered below.

1.3.4 Context Blindness

Vermeulen (2012) specified a feature of WCC, the use of context to create meaning, to be of central importance to neurocognition in autism. Difficulties seen in autism, such as seeing relevance, flexibility in problem solving and understanding language and social cues, rely very much on context (Vermeulen, 2015). Vermeulen (2015) suggested that autistic individuals are capable of seeing the bigger picture but that they do not *use* the bigger picture to create an appropriate understanding of the detail. He considered context blindness as a supplementary explanation of autism cognition rather than a new account (Vermeulen, 2012). The uneven cognitive profile may be explained by context blindness because strengths shown by individuals with autism, such as the embedded figures test, rote memory and rule- and system-based

processes such as mathematics and music, are areas which often require little to no context (Happé and Frith, 2006). There is, however, no current literature suggesting whether context blindness may affect autistic individuals with coexisting intellectual disabilities differently to those without intellectual disabilities and more research is needed in this area.

Context blindness may also go some way to explaining aspects of autism such as sensory processing difficulties which other theories may not account for (Vermeulen, 2015). Vermeulen (2012) suggested that context blindness may result in difficulty selecting the relevant and important sensory input and, therefore, filtering these sensations accordingly. This, in turn, provides explanation for RRBs as discussed in relation to WCC above.

Context blindness also attempts to explain why there may be difficulty ascertaining abilities of individuals on the autism spectrum, namely that assessment conditions in research, diagnostic settings and education are often decontextualised (Vermeulen, 2012). It may, therefore, be a key theory when considering assessment tools and processes for individuals on the autism spectrum with coexisting intellectual disabilities.

1.3.5 Monotropy

One less discussed theory extends the idea of WCC and EF difficulties in autism which involve difficulties in attention shifting. Monotropism attempts to explain why some studies have shown that autistic individuals *are* able to see the big picture in certain tasks (Milton, 2012). Murray et al. (2005) suggested that those on the autism spectrum have highly focused attention which they call monotropism and that their style of attention tends to focus on something to the exclusion of almost everything else. Monotropic tendencies are emphasised as "atypical strategies for the distribution of attention" (Murray et al., 2005, p.139) and are considered a different cognitive style rather than an impairment or deficit. Distributing or shifting attention is necessary for interactions which autistic individuals may find difficult, for example social interaction, use of language, interruptions to structure or routine and unexpected changes. Further, Murray et al. (2005) proposed that individuals with autism may have "no problems in integrating information when it is attended to" but, for this to be the case, the task and goal need to be understood, the process to reach the goal must be known and the individual must have

some motivation to achieve it (p.141). Murray et al. (2005) maintained that, if these points are satisfied, an autistic individual will be able to perform a task successfully.

Monotropism seeks to explain the uneven cognitive profile by suggesting that higher abilities are most likely to exist in areas of interest and, usually, concerning skills which do not involve shifting or distributing attention (Murray et al., 2005). Although it is difficult for individuals on the autism spectrum to distribute their attention in order to make connections across different interests, within an area of interest building connections may not be difficult at all (Murray et al., 2005). This would explain peaks of ability within narrow fields and in-depth knowledge around particular subjects. The hypothesis would need to assume that, for skills shown to be strengths for many with autism such as the block design test or embedded figures tests, these are areas which are of interest to the majority of autistic individuals.

The link between monotropism and restricted interests is evident, with restricted interests due to intense focus and difficulty in distributing attention. Monotropism also addresses repetitive behaviours seen in autism through the idea that monotropism creates "deep basins of attraction where attention gets caught, and may be expressed in a thought or action over and over again" (Murray et al., 2005, p.146). The theory also suggests that when attention is so focused, any interruption can be such an immense and frightening intrusion that it is necessary to repeat the initial action or process in order to return to the "familiar and reassuring" safe state (Murray et al., 2005, p.147).

The difference between autistic individuals with and without intellectual disabilities in relation to this theory is unclear, as those with intellectual disabilities are also likely to have difficulties with attention allocation. Murray et al. (2005) considered, however, that the impact of monotropism will vary between individuals but, for those with additional disabilities, the difficulties in integration will likely be pervasive.

1.3.6 Links and Overlap Between Theories

Much research has attempted to determine the link between these theories, however the relationship between them is yet unclear. Those detailing the theories have sometimes attempted to differentiate between them and, occasionally, to express possible links and overlaps. For example, Baron-Cohen (2009) distinguished E-S from WCC by claiming that systemising is

purposeful and can be overcome whereas WCC is said to be inherent. Frith has previously suggested that ToM deficits are caused by overriding WCC but has since proposed that they may be independent (Happé, 1994). Pellicano et al. (2006) described variations in research: that theories of WCC and EF overlap, that WCC *explains* executive dysfunction and that deficits in EF account for ToM difficulties. Milton (2012) described monotropic focus affecting the ability to make sense of social interactions and therefore resulting in "both apparent and real 'theory of mind' difficulties" (p.7). Vermeulen (2015) maintained that context blindness explains ToM, EF and WCC as the deficits recognised in these theories all relate to context. Further research is needed to determine the nature of context blindness, its relationship to other areas of functioning and its implications in practice.

Research has attempted to find associations between domains described by these theories, however, there have been no consistent findings that any of the domains particularly relate to any other (Happé and Ronald, 2008). Happé and Ronald (2008) reported that, despite some research showing correlation between domains (e.g. Pellicano et al., 2006), research has failed to show clear associations between EF and ToM or between ToM and WCC. Tager-Flusberg (2007) argued that "no single hypothesis can explain the full range of symptoms that defines autism" (p.312). Happé and Ronald (2008) similarly posited that each explanation may independently contribute towards the autism phenotype and that variations in deficits for each individual may determine how autism manifests differently across the population. Perhaps, then, these theories are seeking to explain autism through an equivocal primary deficit while, as Ozonoff et al. (1991) suggested, "instead, it may be more fruitful to consider autism a disorder of multiple primary deficits" (p.1082). Multiple differences, therefore, may not be able to be accounted for by one overarching theory and it is possible, and perhaps even likely, that these theories explain different parts of the cognition of autistic individuals and interact to form a picture of the uneven cognitive profile, communication and language difficulties, RRBs and sensory needs which are characteristic of autism.

1.4 Conclusions

The characteristics, needs and difficulties for individuals on the autism spectrum may be explained, at least in part, by some of the theories described above. The overlap and interaction between many of the cognitive processes explained by these theories reflect a condition which is

particularly heterogeneous in nature. The presence of coexisting intellectual disabilities complicates the picture further still. The implications of this mix of processing strengths and difficulties, along with the presence of intellectual disabilities on skills, behaviours and functional outcomes, results in the need to consider this group separately from those with developmental delay or other developmental disabilities. The functional and behavioural needs described in this chapter and the profile of strengths and weaknesses may also vary from individual to individual within this group and there are implications for our understanding of the likely barriers faced by children with autism and intellectual disabilities within the classroom. The specific needs of this group of pupils cannot be adequately met by generalising education, intervention and assessment to all pupils across the entirety of the autism spectrum or merely considering them within the broader category of pupils with intellectual disabilities. In order to meet their educational needs, the complex nature and interaction between their strengths and weaknesses must be taken into account and the potential resulting difficulties and barriers must be considered when attempting to support individuals within the classroom.

Chapter 2. Educational Assessment of Pupils on the Autism Spectrum with Coexisting Intellectual Disabilities

Autistic pupils with coexisting intellectual disabilities have a unique set of educational needs and potential barriers to learning which have been outlined and discussed in Chapter 1. In order to address these specific needs, it is recognised that the approaches to the education of this group of pupils need to be specific, relevant and, where possible, individualised (Jordan, 2001). However, something which is not often acknowledged and addressed is that the ways of identifying and monitoring skills, abilities and behaviours for this group of pupils also need to account for the unique needs that these pupils have. Recent reviews into aspects of the education and assessment system in England have acknowledged that additional challenges may be present when assessing pupils on the autism spectrum (McIntosh, 2015) but there is no specific guidance on how best to assess this specific group of pupils and no indication of the need for autism-specific assessments which are robust and useful. The current situation regarding assessment in special schools in England[2] will be outlined within this chapter. Types and purposes of school assessment will be described, along with how these are used in special schools. A description of the difficulties of assessing pupils on the autism spectrum with coexisting intellectual disabilities, referring to some of the needs and difficulties discussed in Chapter 1, will be presented. Relevant assessment theories such as psychometric theory and classical test theory will be considered and some of the measurement properties which are necessary to ensure that assessments are robust and fit for purpose will be outlined. Finally, the practical applications of psychological measurement in education, along with potential challenges, will be discussed.

[2] Education is a devolved area for UK governments. This research is mainly concerned with assessment in education as it applies to England although two special schools in Wales were involved in the evaluation stage of this research. Currently, Wales closely follows the English National Curriculum. A new Curriculum for Wales is to be introduced in 2022 and is underpinned by assessment with the publication of statutory guidance encouraging a focus on assessment for formative rather than summative purposes, i.e. to assess learner progress and inform next steps in teaching and learning. In a similar way to England, schools in Wales are also able to develop and implement their own assessment practices alongside statutory assessment.

2.1 Current Situation

2.1.1 Policy and Context

It is necessary to consider assessment within a social and political framework (Delandshere, 2001). Statutory assessment of pupils in special schools in England has changed and evolved in recent years. It was recognised that the more regular use of statutory assessments intended to be for the end of key stages added to teacher workload, led to a curriculum driven and dominated by assessment and had "a profoundly negative impact on teaching" (McIntosh, 2015, p.5). National Curriculum Levels were removed as a statutory assessment for mainstream schools with the introduction of the New National Curriculum in 2014. The statutory assessment system used for pupils working below National Curriculum Level was the Performance Scales (P Scales). The P Scales were widely used in special schools but, after being reviewed in 2015, removal of this assessment was also recommended (Rochford, 2016). The Pre-Key Stage Standards for key stage 1 and 2 were introduced in 2018 and replaced the higher P Scales (P levels 4-8) for pupils engaged in subject-specific study (Standards and Testing Agency, 2020b; Standards and Testing Agency, 2020c). The standards cover English and maths and pupils move onto the Pre-Key Stage Standards at the point that they are working towards targets such as distinguishing between one and lots, showing anticipation about what is going to happen and drawing lines or shapes on paper, in the air or in sand (Standards and Testing Agency, 2020b). For pupils not yet engaged in subject-specific study, the lower P Scales will be replaced by the Engagement Model from the current 2020-21 academic year and will become statutory from September 2021 (Standards and Testing Agency, 2020a). The Engagement Model covers a broader and more general framework of pupil engagement. It is described as flexible and holistic and recognises the nature of non-linear and lateral progress for pupils with SEN. The Engagement Model can be used both summatively and formatively (the definitions of these terms are outlined in detail in the following section), although it is not intended to replace schools' individual assessment systems (Standards and Testing Agency, 2020a). Younger pupils in the Early Years Foundation Stage (EYFS), are assessed against the EYFS Profile which outlines 17 early learning goals and three characteristics of effective learning (Standards and Testing Agency, 2019). The EYFS 2020 Handbook suggests that additional assessment practices may be used for EYFS pupils with SEN such as autism (Standards and Testing Agency, 2019).

For those pupils working below National Curriculum standard in secondary schools or for primary subjects outside of English and maths, there is no statutory assessment specified (Smith et al., 2020). It is necessary to note that all of these statutory assessments are used for pupils with a wide variety of intellectual and developmental disabilities and none are autism-specific.

Aside from the new statutory assessments, schools have been given the freedom to assess pupils in ways which best meet their needs (Office for Standards in Education [Ofsted], 2020) with an aim of reducing teacher workload and shifting focus to teaching and formative assessment (Poet et al., 2018). Schools have specifically been encouraged not to use statutory summative assessments to track and monitor progress between key stages (Standards and Testing Agency, 2020b; Standards and Testing Agency, 2020c). The assessment of skills and progress in areas outside of core subjects, particularly for those with SEN such as autism and intellectual disabilities, varies between schools. Many aspects of the areas discussed in Chapter 1 will not be included in these statutory assessments even though the importance of progress in these broader areas for this group of pupils has been recognised by government commissioned reviews. In the report by the Commission of Assessment Without Levels Report, McIntosh (2015) noted that:

> "Schools should consider meaningful ways of measuring all aspects of progress including communication, social skills, physical development and independence. Assessment should reflect the extent to which a pupil can apply their learning in a wider range of contexts and enable teachers to determine what they need to do to ensure that the intervention and support provided enable children to progress in all areas of their learning and development." (McIntosh, 2015, p.38)

Similarly, when describing the Engagement Model, the Standards and Testing Agency reiterated that additional assessments used in schools should be holistic and personalised to the needs of pupils where possible, and covering areas "related to pupils' early developmental skills, such as their functional abilities in the use of vision, hearing, touch, early communication, behaviour and motor skills" (Standards and Testing Agency, 2020a, p.16). Although the freedom and flexibility for schools to approach assessment in individual ways provides an opportunity for schools, there are some concerns about the burden on individual schools to spend time and resources creating assessment practices that ultimately may end up looking very similar to that developed by other schools (Smith et al., 2020). An additional concern is that schools will need to ensure that

assessments are valid and robust (Smith et al., 2020). Despite these challenges, when considering the implications of the recent assessment changes for autistic pupils with intellectual disabilities who attend special schools, there is an opportunity to use appropriate assessments which reflect the specific needs of this group of pupils discussed in Chapter 1 above.

2.2 Assessment - Types, Purposes and Methods

Theoretically, assessment may be able to be categorised and described. In practice however, assessment occurs in a wide variety of contexts and, particularly when discussing the assessment of pupils in special schools, it is difficult to isolate and disentangle the types, purposes and uses of the assessment in its practical application. Pupils on the autism spectrum who are educated in special schools are likely to undergo assessment for a variety of purposes. This section will briefly describe and differentiate between different types and purposes of assessment, consider the uses of the different assessments in special schools and discuss the implications of their uses for pupils on the autism spectrum with intellectual disabilities.

2.2.1 Types and Purposes of Assessment

2.2.1.1 Diagnostic and Screening Assessment

Clinical screening and diagnostic assessments are almost certain to have been undertaken in order for the individual to have received a diagnosis, to have the extent of their educational needs determined and for an appropriate educational placement to have been obtained. Diagnostic assessment may be referred to by schools to gauge a pupil's baseline skill prior to any teaching or instruction (Lewis et al., 2003) and screening assessments may be used order to determine interventions and teaching strategies which are appropriate for an individual and to adapt teaching accordingly (Charman et al., 2011a). Gipps (1994) also identified that diagnostic assessment can be conducted informally by a teacher or those working with the child to identify problems or difficulties in learning. This research is not concerned with diagnostic or screening assessment, however it is necessary to note that, for pupils with SEN particularly, sometimes assessments are created for multiple purposes and the distinction between screening and outcome assessments might not always be clear.

2.2.1.2 Summative and Formative Assessment

Assessments used in schools are considered to be either summative or formative. Summative assessment is usually a more formal, and often statutory, method of assessment which records a pupil's level of achievement at a particular point in time. In mainstream education, this is usually conducted systematically in the form of end of module tests or assignments (Harlen and James, 1997). In special schools, however, summative assessment is often recorded through teacher reporting, such as end of year reports, annual reviews and portfolios which consider pupil achievements against statutory assessment criteria. The purposes of summative assessment include to show pupils' abilities and progress, to identify pupils' strengths and weaknesses and for school and teacher accountability (Delandshere, 2001).

Formative assessment, although vaguely defined and difficult to evaluate (Dunn and Mulvenon, 2009), is usually informal assessment conducted by teachers and takes place regularly throughout the learning process. Formative assessment assists learning by allowing teachers to recognise what the pupils can do, identify gaps in their pupils' knowledge and informing next steps for teaching. Formative assessment has been promoted in the United Kingdom since the introduction of the National Curriculum in 1988 (Wilkinson and Twist, 2010) and was recognised in education as a result of the rise of socio-constructivist theories of learning (Baird et al., 2017). It is considered a "practice-based approach to assessment" but the effects of this kind of assessment are rarely evaluated in a formal or systematic way (Baird et al., 2017, p.24). When considering special needs teachers' views and use of formative assessment, Rouse and Agbenu (1998) found that teachers considered informal formative assessment to be mainly about knowing their pupils and they usually considered this to be done intuitively. Similarly, McNicholas (2000) found that special needs teachers of pupils with profound and multiple learning difficulties did not always record formative assessment and, where they did, recording varied greatly between schools and teachers. The difficulties with lack of recording is that teachers' formative assessments may be subject to bias (Baird et al., 2017). Harlen and James (1997) discussed the nature of formative assessment and concluded that it is likely to always be incomplete, as a pupil's ability will vary depending on many factors. However, they also noted that effective assessment is unlikely to ever be exact and precise without potentially harming the learning in schools (Harlen and James, 1997).

Although summative and formative assessment are generally considered to be separate entities, the distinction between the two is still not overly clear and this is often reflected in practice (Baird et al. 2017; Harlen and James, 1997). Rouse and Agbenu (1998) recognised that one form of assessment leads to another and said of formative assessment:

> "its intended purpose is to inform curriculum and lesson planning, to help with the setting of objectives, to provide evidence of pupil progress *and to help with a summative judgement at the end of each phase of education*" (Rouse and Agbenu, 1998, p.82).

Dunn and Mulvenon (2009) acknowledged that although assessments may be "designed and packaged as a formative and summative assessment", it is the implementation and use which will determine which it becomes (p.2). In special education the delineation between the two forms of assessment may be considered even less evident. Where the needs and levels of pupils mean that national mainstream summative assessments such as Standards Assessment Tests (SATs) are not relevant, less formal assessment methods may form the summative assessment. Portfolios of pupil work and achievements, including photographs and staff annotations for example, may be collected as formative assessment and then used as summative assessment to evidence achievement against statutory assessment criteria (Harlen and James, 1997). As previously mentioned, current education and assessment policy appears to be discouraging the use of statutory summative assessment as formative assessment, even though the new engagement profile explicitly states that it can be used both summatively and formatively (Standards and Testing Agency, 2020a). Summative assessments may be used in this way due to a lack of clear and relevant alternative assessments and, as discussed, the distinction between the two different types of assessment may not be straightforward in practice.

2.2.1.3 Norm-Referenced, Criterion-Referenced and Ipsative Assessment

Norm-referenced assessment considers an individual's skill and abilities in relation to their peers (Gipps, 1994). Norm-referenced tests may provide comparative information on a pupil's ability but give little to no information on how to address any difficulties they may have (Harlen and James, 1997). Educational assessment has moved away from norm-referenced assessments in more recent years, although norm-referenced assessments may still have their

place in education when looking to determine the additional needs a pupil may have (Gipps, 1994). While in some respects it may be necessary for norms to be taken into account, good progress or outcomes for special school pupils are likely to differ from mainstream education and, for pupils with autism and intellectual disabilities, will involve many of the areas considered in Chapter 1 above. Norm-referenced assessments using a mainstream population may be of little use to teachers and parents when measuring these small steps of progress. As Wolf-Schein (1998) suggested, if a child is assessed using an assessment which was normed on a different population, a norm sample of younger children for example, "the use of the norms of the test as an index for evaluating that child's current performance or for predicting future performances may be inappropriate" (p.42). Assessments which are normed on a specific population in question may sometimes be useful for comparative purposes, however given the heterogeneous manifestation of autism along with the variations in cognitive profiles discussed previously, particularly when considering the interaction of coexisting intellectual disabilities, comparison between pupils in special schools is not usually a main concern for teachers or parents. Cronbach (1970) stated that "the most useful norms permit the tester to compare the subject with his prospective companions and competitors" (p.106). Although perhaps semantic to some degree, with the focus of special schools on individualised and person-centred targets and progress, it is unlikely that special school teachers' or parents' main concerns are ever with these types of comparisons between pupils. Norms may be useful to identify that a pupil has difficulties but not so useful to measure outcomes when their abilities are either significantly lower than their typically developing peers or vary greatly in different areas even in comparison to pupils with similar needs.

The move in educational assessment towards criterion-referenced assessment is concerned with measuring the level of understanding or ability that an individual has achieved against a standard or criteria (Davis, 1998). Criterion-referenced assessment is used across the whole education system and since the 1980s almost all assessments used in U.K. education are criterion-referenced (Gipps, 1994). Although the individual nature of targets for pupils in special schools means that the need to assess 'mastery' or attainment of a particular level is not perhaps as relevant as it may be in mainstream education, criterion-referenced assessments are important for particular skills that a pupil might need to achieve before others and may also be helpful for teachers and schools to assess key skills which will be worked towards over time. Although often

thought of as a summative assessment due to the presence of criteria, this type of assessment was developed to measure abilities along an achievement continuum and "not necessarily [to] refer to final end-of-course behaviour" (Glaser, 1963, p.519). In this way, criterion-referenced assessment could be used as either formative or summative assessment.

The concept of ipsative assessment, although not widely referred to in education, may be particularly appropriate for pupils on the autism spectrum with coexisting intellectual disabilities. Ipsative assessment is defined as that which compares to an individual's previous performance. It lacks a competitive element, has individualised goals and takes account of the starting position of the individual (Hughes, 2014). Hughes (2014) recognised that, for individuals who may not be considered successful by the standards of norm- or criterion-referenced assessment, ipsative assessment is a way in which their progress and achievements can be assessed, identified and celebrated. Forms of ipsative assessment are already apparent in education and are often combined with formative teacher assessment and criterion-referenced assessment, for example where there is a focus on individual progress with an overall aim of meeting particular criteria or standards. This approach to assessment seems appropriate for the assessment for pupils with intellectual disabilities where it is important to measure whether progress is occurring but the rate or level of attainment may not be relevant in comparison to norms of the whole population or their peers. The concept of ipsative assessment also fits with the very personal targets for pupils in special education and the often individualised routes to these targets. Ipsative assessment is likely to be used intuitively within special schools, especially in relation to informal formative assessment where teachers know their pupils well and are continuously monitoring progress that pupils are making. This holistic approach to assessment and progress also relates to the literature on person-centred approaches to support children and adults with autism throughout their lives (e.g. Beadle-Brown et al., 2009).

As mentioned, there may be overlap in these types of assessment, particularly in special education when the differences between and changes in skills and abilities may be small. Summative assessments may be used formatively and criterion-referenced assessments may be normed on certain populations and used in an ipsative manner. As recommended by Hughes (2014), a mixture or hybrid of assessment types is likely to be appropriate for use in special education to gain an overall picture of the progress and achievements of individual pupils.

2.2.2 Methods of Assessment

There are a variety of assessment methods used in special education. The cognitive levels of pupils at special schools often preclude the use of traditional paper and pencil tests or work to be assessed, particularly for the younger pupils and, therefore, different methods of assessment to those used in mainstream education may be needed (Wilkinson and Twist, 2010).

2.2.2.1 Performance Assessment and Observation

Direct assessment involves evaluation of a final product or of an individual completing a task. Performance assessment refers to the application of the actual skills to show how a pupil can use what they have learnt (Gipps, 1994). Authentic assessment is described by Gipps (1994) as a type of performance assessment where the task is performed in as true a context as is possible. She recognised that assessment can vary in the way it is authentic including aspects of "the stimulus; task complexity; locus of control; motivation; spontaneity; resources; conditions; criteria; standards and consequences" (p.99). Authentic assessments are important in special education to ensure that pupils have functional applications of the skills they have learnt. This may also be particularly appropriate for pupils on the autism spectrum to gauge an accurate picture of skills and abilities in light of the theory of monotropism described in Chapter 1. Schools may be limited in their ability to assess performance however, especially when teaching skills in a specific environment which then require generalisation into other settings or contexts.

Teaching necessarily involves a great deal of observation of pupils by teachers and observation can make up both formative and summative assessment. When no tangible work such as an essay or a completed construction is produced, then observation is the obvious method of assessing pupils' abilities. Depending on the pupils' skills, observational assessment, particularly informal observation, is one of the more common forms of assessment used in special schools. Observational assessments may be structured, where a situation or context is constructed in order for the pupil to be observed performing certain tasks or behaviours, or naturalistic, where an individual is being observed in a more natural environment or context. It is also possible for assessments to be semi-structured, with an example of a semi-structured observational assessment in a clinical, diagnostic context being the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000). Observational assessments may or may not

require instruction, direction or participation from an adult. When discussing these types of assessments with children with intellectual disabilities, Wolf-Schein (1998) termed these variations intrusive and non-intrusive assessments. Intrusive assessments require the individual to engage or interact with an unfamiliar examiner which, she recognised, may result in a performance below what the child is capable of or, in some situations, even in the child being labelled "untestable" (p.36). Similarly, Westby et al. (1996) cited Bailey and Wolery (1989) who identified the limited usefulness of assessments if conducted by unfamiliar people in unknown contexts. A better option, Wolf-Schein (1998) argued, is nonintrusive assessment where the individual being assessed performs within familiar environment, context and routines and does not involve constraints on time or exposure to the unfamiliar. She recognised that, if necessary, interaction should be with a familiar individual who knows the child well to give input on how the child best understands or responds.

The features of observational assessment are relevant depending on the purposes of the assessment. If the specific skills to be assessed include the ability for a child to generalise their abilities to less or unfamiliar adults or new environments, then it is necessary for this to be reflected in the assessment. If the assessment is merely to ascertain how independently or to what extent a pupil is able to complete a task, the assessment that most appropriately reflects a pupil's ability is likely to involve the context in which the activity is naturally going to be carried out. In this way, it is not possible to claim that one version of observational assessment is superior to another until the specific purpose of an assessment has been determined. For observation to be most effective, however, it can be time consuming and may require an objective observer to assess predetermined skills using specific and detailed recording (Merrell, 2001) which, although appropriate in clinical assessment, may prove problematic in school environments.

2.2.2.2 Checklists and Rating Scales

Checklist and rating scales completed by teachers or parents are a further method of assessment often used in special education. Checklists and rating scales can be used as tools in direct observation or as indirect respondent report measures. Checklists will usually include dichotomous responses where skills are marked 'yes/no' or 'present/absent' whereas rating scales allow the informant to consider to what degree certain skills are present or how often they are displayed (Banerjee, 2016). Checklists and rating scales can also be self-report measures,

although this may or may not be appropriate for pupils in special schools dependent on their abilities. Checklists and rating scales are devised to be less labour intensive than observations. One advantage of these types of assessment compared to observation and performance assessment is that respondent reporting allows for a more rounded assessment of a pupil as they can utilise a familiar person's knowledge of a pupil and their abilities over time rather than during a specific observation where a pupil's true or usual capabilities may not be shown (Merrell, 2001). Current and former statutory assessments such as the P Scales, Pre-Key Stage Standards and the Engagement Model are all reflective of this method of assessment, allowing teachers to report on pupils' abilities for each descriptor or standard with supporting evidence provided to back up and justify decisions. As Merrell (2001) suggested, continuing improvements in the development and use of rating scales and the evidence supporting their use means that they are now widely used to assess children's abilities and behaviour. Teacher bias may be a potential problem in using teacher ratings, particularly if assessment is used summatively for quality control on teaching and it is necessary to ensure that rating scales and checklists in schools have been evaluated for sufficient reliability and validity to reduce the likelihood of inaccurate assessment. However, if well developed and evaluated to ensure they are robust, checklists and rating scales are convenient assessment methods which may allow for a rounded view of a pupil's abilities, skills and behaviours.

2.2.3 Summary

There are a variety of different types, methods and purposes of assessment used in education. Assessment processes in special schools necessarily differ to that of mainstream schools due to the pupils' ability levels. When considering the practicalities of assessment in special schools, the distinction between the different assessment types, formative and summative assessment for example, may be somewhat blurred compared to mainstream school assessments. Similarly, it may be more usual for assessments to be used in different ways for different purposes than in mainstream schools. An important aspect of the assessments, especially in special schools where the delineation between types of assessment is not clear, is how they are used. Cronbach acknowledged in 1970 that "tests must be selected for the purpose and situation for which they are to be used" and this is still incredibly relevant today (Cronbach, 1970, p.115). If the purpose of the assessment is to capture an individual's abilities at that time, compare them

to others, or hold teachers and schools to account, then norm-referenced, standardised, summative assessment is likely to be most appropriate. If the aim is to monitor progression and improvement in certain areas of learning or abilities, formative assessment with an ipsative approach may be more useful. Again, there may still be some overlap in these types of assessment but, crucially, the assessment must make use of appropriate assessment methods and be matched to the purpose to ensure that the abilities of pupils are accurately identified and that progress towards main learning goals can be shown.

2.3 Difficulties Assessing Autistic Pupils with Coexisting Intellectual Disabilities

The assessment of those with coexisting autism and intellectual disabilities in special schools may be complicated by a number of difficulties. The assessment of internal systems such as cognitive processes, thoughts and behaviours is a difficult task. As Pellegrino (2014) recognised, "assessment results are only estimates of what a person knows and can do" (p.68). These difficulties are compounded when attempting assessment of individuals with additional impairments in communication and cognition and the most notable difficulties in assessing this population relate to the practicalities of assessment. Some of these difficulties have been acknowledged in the DSM-V which states that the "assessment of intellectual ability may be complicated by social-communication and behavior deficits inherent to autism spectrum disorder, which may interfere with understanding and complying with test procedures" (APA, 2013, p.40). The practical difficulties of assessing those on the autism spectrum with coexisting intellectual disabilities may be two-fold; difficulties due to social communication impairments and difficulties in understanding, as well as difficulties due to behavioural issues, lack of motivation or non-compliance. Only relatively recently has there been a change from viewing individuals with more severe intellectual disabilities as "untestable" (Wolf-Schein, 1998, p.36) to adapting assessments to make them appropriate for these populations. The impact of these difficulties is still evident in research where individuals with the most severe and complex disabilities are often excluded due to the additional complication assessing those with intellectual disabilities (Brugha et al., 2015).

2.3.1 Communication

Wolf-Schein (1998) wrote in detail about the difficulties of assessing children on the autism spectrum in the late 1990s. She described assessment and communication as being "inextricably linked" and identified that "the better the ability of the child to communicate, the easier it will be to ascertain his or her competence" (Wolf-Schein, 1998, p.35). The assessment of a minimally or pre-verbal child or those with limited receptive understanding may prove particularly difficult. Effective reading assessments of minimally or pre-verbal children, for example, need to look very different to traditional reading assessments requiring verbal output (Arnold and Reed, 2016). The need for specific assessment tools relevant to those with more severe intellectual disabilities has more recently been accepted in healthcare research. Recent interest in health screening for individuals with intellectual disabilities has recognised that the use of specific assessment tools for these individuals can result in beneficial health outcomes (Barney et al., 2020). When considering the social, communication and language impairments central to an autism diagnosis and combining those factors with intellectual disabilities, the picture may be complicated further. The difficulties assessing those on the autism spectrum using educational assessments may also be reflected by the fact that intelligence testing and diagnostic tests for medical or mental health conditions, even when they are specific to those with intellectual disabilities, may prove less effective for autistic individuals. Similarly, Happé (1994) questioned whether IQ tests can be considered valid for those on the autism spectrum due to the uneven cognitive profile, as skills tested may be specific and not a valid assessment of a whole domain. Brugha et al. (2015) discussed similar difficulties with mental health assessments and stated that many mental health self-report measures used in diagnostic screening may not be appropriate for autistic individuals and those with intellectual disabilities due, not only to their difficulties with communication, but also to differences in interpretations of emotion and abstract thinking. This further supports the view that it may not always be enough to use assessments developed for those with special needs or intellectual disabilities; the specific and particular needs of individuals on the autism spectrum may require assessments which have been developed to take their needs into account and this is also relevant to assessment in education.

2.3.2 Motivation and Compliance

A further practical difficulty in assessing those on the autism spectrum with coexisting intellectual disabilities is that of motivation and compliance. Again, Wolf-Schein (1998) identified that assessments usually "assume a cooperative child who will be trying to achieve and whose responses will reflect what they are most capable of doing because they understand the situation and the task and are willing to try their best" (p.35). Lack of compliance or lack of motivation may result in an inaccurate reflection of skills and abilities. RRBs, for example, may affect motivation and mean that a task presented in different ways may result in different responses. A pupil may readily complete a task which they are interested in or which contains elements of special interests but may not engage with a similar task which holds no element of interest or for which they don't see a purpose. Similarly, pupils may not engage with a task they find too simple and, therefore, linear assessments, where an earlier level must be completed before moving on, may result in a pupil's abilities being underestimated. This links to the theory of Monotropy discussed in Chapter 1. Murray et al. (2005) described four requirements for an individual to perform a task: seeing the purpose of the task, valuing that purpose, understanding what is required and knowing the steps to achieve it. They proposed that "it is important to make sure that any testing of individuals with autism spectrum disorders meets these requirements, or it may not measure what it purports to measure" (Murray et al., 2005, p.141). In other words, it is important to ensure that a test of ability in a specific area, domain or of a particular skill is genuinely assessing the relevant skill rather than an individual's ability to focus or understand the instructions, for example. It may be that these difficulties are apparent when assessing any young child, however the key characteristics of autism, such as communication difficulties and RRBs discussed above, will no doubt exacerbate these difficulties. The theory of Context Blindness may also account for challenges with assessing those on the autism spectrum due to not using the bigger picture in order to create an appropriate understanding of detail (Vermeulen 2015). Context Blindness may explain potential difficulties in ascertaining abilities of autistic individuals, namely that assessment conditions in research, diagnostic settings and education are often decontextualized (Vermeulen, 2012). If the task itself and the context of the task can be personalised to ensure that it is relevant to the individual pupil, the outcome of the assessment may be more accurate. Similarly, the assessment of whether pupils have generalised skills to appropriate settings is likely to be difficult as the possibility of contextual performance or

authentic assessment may be limited in a school setting. It may be more appropriate to conduct assessment using a respondent, such as a teacher or parent, where knowledge of the pupil and their abilities by someone who knows them well is taken into account to provide a more holistic account of their abilities.

2.3.3 Content of the Assessment

Along with the need for assessments to be appropriate for both the population they are assessing and the assessment purpose, the content of the assessment is also of particular relevance in a number of ways.

Firstly, it is important that relevant and appropriate skills, abilities and behaviours are being recorded. This relates to the face and content validity of an assessment; ensuring that it covers the appropriate aspects of a domain and relevant specific skills and behaviours which it is required to assess. Content validity is usually evaluated through use of experts in the relevant area. In educational assessment, it would be appropriate to consider the teachers to be experts in the content they are teaching and, as they are the administrators, interpreters and scorers of the assessment, that they would be consulted on the content of assessments which they will be using. There has been a suggestion by teachers in recent years, however, that the content of assessments is not always appropriate and this would suggest that teachers are not always being utilised for content validity evaluations of teacher and school assessments. Small improvements of skills described in Chapter 1 may represent large and important achievements for autistic children with severe to profound intellectual disabilities (Pellicano et al., 2014) but, as mentioned, many of these basic skills are not part of the focus of statutory assessment (Standards and Testing Agency, 2020a; Standards and Testing Agency, 2020b; Standards and Testing Agency, 2020c). Content of assessments has been questioned further, with some commentators asking whether assessments are even "capable of measuring the things that will make the biggest difference in children's lives" and that question is one which has been echoing in mainstream and special education circles in recent years (Resnick, 2017, p.151; Baird et al. 2017). Parent and professional groups have suggested that a broad repertoire of behaviours should be included when measuring progress and outcomes such as social interaction skills, everyday adaptive skills and recognition of difficulties (McConachie et al. 2018). Good short- and long- term outcomes will be highly individualised for each child and their family and may not equate with the

traditional definition of achieving independence and a social life comparable to that of typically developing individuals (Ruble and Dalrymple, 1996). The idea that assessments "end up measuring things that they can measure more easily" rather than areas that are considered priorities by teachers and parents, again, relates to the content validity of assessments and how this is addressed and evaluated (Resnick, 2017, p.151). This is also reflected in areas of psychology where commentators have recognised that "it is better to imperfectly measure relevant dimensions than perfectly measure irrelevant ones" (Bommer et al., 1995, p.602). It is particularly important that relevant skills, abilities and behaviours are assessed for the purposes of the assessment and that content validity is evaluated effectively.

Secondly, in order to show progress, it is crucial that the assessment will pick up the appropriate amount of change. For pupils on the autism spectrum with intellectual disabilities, for example, small steps in progressing towards a particular goal may be important (Pellicano et al., 2014), and therefore an assessment may need to pick up on smaller amounts of change compared to an assessment in mainstream schools where learning may happen more quickly. This may mean that an assessment needs to be able to identify and address these small steps in skills, abilities or behaviours or perhaps have some flexibility in the way these are accounted for. This links to the responsiveness of a measure, which is its ability to detect change over time (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018).

Finally, even if assessments are developed and evaluated specifically with individuals with autism and intellectual disabilities, the heterogeneous nature of autism may still prove problematic for assessment of this population (Rivard et al., 2015). With the variation of manifestation of autistic characteristics and behaviours, it is difficult for an assessment to accurately allow for the various ways that an individual may show a skill or ability if the content is too specific. One example of difficulties associated with tightly defined content is a type of criterion-referenced assessment called domain-referenced assessment. It uses "formal sampling rules" and is often focused on very specific areas which Gipps (1994) argued "leads inevitably to reductionism, where the content is specified more and more minutely and attention is diverted away from broader achievements" (p.82). She suggested that, although this criticism is often conflated with criterion-referenced assessment, the latter can be broad and does not need to be so focused and detailed. Domain-referenced assessments may be relevant for use in special schools,

addressing specific domains of difficulty or using the smaller, more tightly focused areas to show smaller steps of progress towards larger goals. The criticism that domain-referenced assessments may neglect broader achievements is valid, however, and the danger is that achievements outside of the narrowly defined goals may not be recognised.

There is a need for balance in educational assessments used in special schools between the content being relevant and specific enough to show small steps of progress and the need for an assessment to be holistic and flexible when reflecting the needs of this heterogenous group of pupils.

2.3.4 Summary

Gaining an accurate picture of the abilities and skills of individuals with intellectual disabilities through assessment is evidently more difficult than assessing pupils in mainstream schools. The interaction between autism and intellectual disabilities may result in additional difficulties in accurately assessing the skills, abilities and behaviours of pupils on the autism spectrum. Complex communication difficulties and perception of context resulting in lack of motivation may be factors which add to the difficulty of assessing this group of pupils. Assessment content is also relevant and it is necessary to ensure that the content of an assessment appropriately reflects the areas being assessed as well as accounting for the heterogeneous nature of the condition where possible. Despite the specific needs for this group of pupils and the potential impact upon assessment, which has been recognised in government commissioned reviews and reports, few assessments specific to this group are in common use in schools (McIntosh, 2015). Routinely, pupils on the autism spectrum in special schools are assessed using generic assessments which were developed for pupils with intellectual disabilities or developmental disabilities regardless of specific diagnosis (Arnold and Reed 2016). It is important that teachers in special schools have appropriate ways of assessing and showing progress for autistic pupils with assessments which are designed with the needs of these pupils in mind and account for the difficulties in assessing their skills, abilities and behaviour.

2.4 Assessment Development and Evaluation

Accurately measuring any aspect of ‘the mind’ can be challenging and, as discussed above, when it comes to pupils with additional and complex learning needs, there may be

confounds which further exacerbate these difficulties. Although in education the focus is usually on what a pupil can do, just because an autistic pupil with intellectual disabilities does not show that they can perform a task or display a skill, does not necessarily mean that they cannot. The understanding of and motivation behind the need to 'perform' may not be the same as for pupils in mainstream school. Similarly, difficulties with interpreting context and the consequent challenges of generalisation may result in assessments not accurately reflecting a pupil's abilities in different situations or settings. The initial, seemingly simple consideration of directly determining what a pupil can do is complicated by the difference between what they truly can do and what they may show they can do. Even when considering easily observed skills, abilities and behaviours, the areas and domains become more abstract and conceptual when the pupil's true abilities are the focus. These are all factors that need to be taken into consideration when developing or evaluating an assessment alongside the appropriate literature on measurement, assessment development and evaluation which will be considered in the following sections.

2.4.1 Psychometrics, Classical Test Theory and Measurement Properties

Measurement is the process of systematically assigning a number, value or quantity in order to describe or represent the features or properties of objects, events or individuals (Raykov and Marcoulides, 2011). Psychological measurement, then, is the process of quantifying traits and qualities of an individual or group of individuals. Although these abstract and unobservable traits and qualities, known as constructs, cannot be directly seen, certain aspects of behaviour that are observable may indicate the existence and extent of these traits (Raykov and Marcoulides, 2011). For example, regulation of emotion is abstract, intangible and cannot be observed. However, an individual's behaviour in certain circumstances may indicate that they can regulate their emotions and these behaviours may be able to be assessed or measured to give an indication of emotional regulation. In psychological measurement, the existence and extent of the construct is indicated by the directly observable behavioural manifestations of the construct (Raykov and Marcoulides, 2011). To some extent, almost all educational assessment is psychological, as an individual's knowledge of a subject cannot be observed but they can write, explain or perform in ways which mean their knowledge can be assessed.

Psychometrics is the field concerned with psychological measurement and acknowledges that all assessments are imprecise and subject to error. Psychometrics provides a way to

determine to what extent this is true of particular assessments. The idea that there may be disparity between what an individual can do and what can be measured is reflected by Classical Test Theory (CTT). CTT suggests that no psychological measurement is without error and that a person's observed test score is the sum of their true score and the error score. The application of this main concept of CTT makes sense in relation to the assessment of pupils on the autism spectrum with intellectual disabilities in special schools; the difference between a pupil's true ability and what is observed or recorded is the error of the measurement. As discussed above, there are many difficulties in assessing the abilities of this group of pupils which may be due to a variety of reasons including the pupil not understanding what is asked of them, not showing what they are capable of, or attributable to the person reporting or recording. Measurement error can be systematic or random; systematic measurement error is consistent or regular without relevance to the construct, while random measurement error is unrepeatable or due to chance effects unrelated to the construct (Raykov and Marcoulides, 2011). As Raykov and Marcoulides (2011) noted systematic error cannot be present if a measurement is to be considered trustworthy. Unlike systematic error, the mean average of truly random error over multiple administrations of a test is zero.

Psychometric validation focuses on the evaluation of the measurement properties of assessments. It uses statistical and mathematical methods and models to overcome or minimise the difficulties and challenges associated with psychological measurement (Raykov and Marcoulides, 2011). Measurement properties of outcome assessments include various aspects of reliability, validity and responsiveness. Reliability is concerned with measurement error as outlined by CTT. Lack of measurement error would equate to perfect reliability, therefore the more measurement error that exists, the lower the reliability of the measure (Leppink and Pérez-Fuster, 2017). The reliability index answers the question of how much the observed scores relate to the true scores and is defined as the ratio of the true score standard deviation to observed score standard deviation (Raykov and Marcoulides, 2011). The reliability coefficient is equal to the squared reliability index and is defined as the ratio of true score variance to the observed score variance (Raykov and Marcoulides, 2011) or the "proportion of observed score variation which is not attributable to measurement error" (Johnson and Johnson, 2009, p.11). Procedures to estimate reliability may include the split half method, alternative forms, test-retest reliability, inter-rater reliability (IRR) and internal consistency reliability (Raykov and Marcoulides, 2011).

Specific methods used to evaluate the measurement properties may vary due to the nature of the assessment or scale, and the methods relevant to this research will be outlined further in the description of scale development in Chapter 3.

Reliability is a necessary but not sufficient condition of validity. Validity refers to whether an assessment is measuring the construct that it was developed to measure. There are various aspects to consider and methods to provide evidence for the existence of validity. However, because validity is a question of degree, it cannot be suggested with absolute certainty that a measure is completely valid (Raykov and Marcoulides, 2011). It is also necessary to note that validity is contextual and, as has been discussed previously, evaluations may suggest that a measure is valid in one setting or with one population but may not be valid in others (Cronbach, 1970). There are various facets of validity which can be evaluated to lend support towards a conclusion of the validity of a measure as a whole. Face validity is an evaluation of whether the content of a measure appears to cover the relevant areas whilst content validity is a deeper analysis of whether the assessment content appropriately represents the construct as it is defined (Raykov and Marcoulides, 2011). Criterion validity refers to how scores on one measure predict or correlate with scores on another measure. This can be in relation to how an individual will perform in the future (predictive validity), or at the same time on a different assessment (concurrent validity). Concurrent validity can be further broken down to when scores on the measure are expected to correlate with scores on another assessment (convergent validity) or not (discriminant validity). Hypotheses may relate to strength and direction of correlations between certain measures or predict no correlation between two measures. Although all of these aspects of validity point towards the construct validity of an assessment, further evidence of construct validity can be determined through methods such as factor analysis (Raykov and Marcoulides, 2011). This is sometimes referred to as structural validity as it is concerned with the internal structure and dimensionality of a measure, i.e. whether the assessment is unidimensional and measures just one construct or, if not, how many dimensions or factors exist in a measure and which items load onto which factors or constructs. Face and content validity are usually associated with assessment development, with other aspects of validity and reliability relevant to validation or evaluation procedures. Further aspects of validity such as cross-cultural validity, group comparisons, sensitivity and specificity may be relevant to certain types of assessments (e.g. translated measures or diagnostic and screening assessments) but may not be necessary to

consider when evaluating outcome measures in schools. As noted above, the methods used to evidence the appropriate aspects of validity will be outlined further in Chapter 3.

Responsiveness is the ability for the assessment to measure change in the relevant construct over time (Prinsen et al. 2018; Mokkink et al., 2018; Terwee et al., 2018). Responsiveness is a measurement property of particular relevance to outcome measures and addresses the validity of the change score of a measure (Mokkink et al., 2010). Responsiveness can be evaluated post-intervention using effect size to test a hypothesis on the amount of change that has occurred or can be determined through comparison with the responsiveness of another measure.

Many of these aspects of validity and reliability are relevant to educational assessment depending on the assessment type, purpose and context. Where the focus is on showing what autistic pupils in special schools can do and identifying and recording progress in their potential barriers to learning described in Chapter 1 above, specific measurement properties will be particularly relevant for evaluation. To ensure that an educational assessment has content validity, it is important to engage experts, in this instance the teachers, in the construct definition and the assessment content (Baird et al., 2017). Depending on the specific purposes of the assessment, criterion validity is likely to be more relevant to progress assessments in special schools than predictive validity and, similarly, convergent validity may be more important than discriminant validity. Further evaluation of structural validity such as factor analysis may be appropriate depending on the assessment model (Prinsen et al. 2018; Mokkink et al., 2018; Terwee et al., 2018). Relevant aspects of reliability include the reliability of the assessment over time (test-retest reliability), the reliability of the assessment items (internal consistency reliability) and the reliability of the assessment judges, raters or markers (IRR). Evaluating the responsiveness of a measure to determine how well it captures change is a particularly important measurement when using assessments to monitor progress in special schools.

Two alternative models in the analysis of measurement, which are extensions of CTT, include Item Response Theory (IRT) and Generalizability Theory (G Theory). IRT, a complementary rather than comprehensively superior approach, conducts analysis at item level whereas CTT analysis is performed on the test as a whole, with any item level analysis applicable only to the specific circumstances of that analysis (Wu et al., 2016). IRT, sometimes

referred to as modern psychometrics, is widely used due to advantages such as the lack of dependence on sample for test characteristics and the ability to predict responses for items depending on an individual's latent traits (Zanon et al., 2016). In respect of reliability evaluations, G Theory goes further than CTT in that it isolates different sources of measurement error, whereas measurement error in CTT is undifferentiated (Vispoel et al., 2018). Psychometric analysis using IRT and G Theory were not included in this research, firstly as CTT was sufficient for the basic preliminary evaluation presented in this thesis (Wu et al., 2016) and, additionally, due to the large sample size required by these approaches (Atilgan, 2013; Brown and Abdulnabi, 2017). Further evaluation using these theoretical approaches is needed within future research.

2.4.2 Assessment in Psychology and Education

There are a number of practical and theoretical differences between assessment in psychology and education (Pellegrino, 2017). Traditionally, psychology and psychometric approaches to measurement have been concerned with between-subject differences. In education, although norms and cohorts are relevant in some contexts, much of the focus currently is on individual change, particularly when it comes to educational assessment. The differences in approach to assessment between psychology and education have often been argued on a theoretical basis (Tennent, 2020). There is disagreement, for example, between the traditional, critical-realist approach of psychological assessment, which considers validity as being specific to the assessment content, and the more constructivist educational stance, which places high value on the assessment purpose when considering its validity (Tennent, 2020). Psychometricians may argue that the purpose of assessments *is* taken into account during the validation process, one example being validation with different and specific populations. This consideration of the purpose of an assessment was also highlighted by Cronbach (1970) who stated:

> "We cannot ask the general question 'Is this a valid test?' The question to ask is 'How valid is this test for the decision I wish to make?' or 'How valid is the interpretation I propose for the test?'" (Cronbach, 1970, p.122)

The disparity, then, may be the *extent* to which context around the assessment is accounted for and the value that is placed on either the statistical psychometric validation or the practical

usefulness of an assessment. Baird et al. (2017) suggested that, where there is conflict between educational objectives and measurement, education should take priority. They posited that resolving any differences in perspective is likely to require collaboration between the fields while also recognising that priorities are always liable to remain different (Baird et al., 2017). For an assessment to be useful yet still stand up to scrutiny, a path must be navigated between two approaches, each concerned with their own frame of reference and with the priorities of each perspective potentially masking the advantages of the opposing point of view.

When considering special needs education, as with many other types and purposes of assessment, the line between psychological and educational assessment may begin to blur. Many assessments used in clinical screening or identification address areas relevant to skills and behaviours which are focused upon in schools, for example adaptive behaviour, communication and behaviour that challenges. This is recognised by the fact that some assessments have had separate school versions of their measure created for and validated in school settings. The use of school versions of psychological assessments covering relevant areas may clearly be appropriate for use by teachers in special schools in certain circumstances. However, a test can be found to be psychometrically sound but if it is not appropriate for or useful in a particular situation, setting or context, then ultimately that test will not be valuable to teachers for that purpose. There are not only theoretical differences between the two fields, but also a more general gap which often exists between theory and practice, particularly in education (Vanderlinde and van Braak, 2010). As Johnson and Johnson (2009) stated in a report commissioned by the Office of Qualifications and Examination Regulations, there is a need to explore issues of assessment reliability, for example, against a "complex reality" (p.2). Psychometric evaluation needs to be balanced against the usefulness of a test in an education setting to ensure that an assessment is robust but also that it is appropriate and valuable to teachers for the purposes for which it is to be used.

Education as a field is still in its infancy compared to disciplines such as mathematics and the sciences and is also more likely to be influenced and driven by policy (Baird et al., 2017). Although routes forward for psychological and educational assessment are being discussed (Wiliam, 2017), with a lack of consensus on robust validation procedures for educational assessment, and considering the links and overlaps discussed above, this research will use

psychometrics as the basis for evaluating assessments. In order to balance psychometric validation with the need to ensure that this research remains relevant to education and practice in special schools, the context of validation will be crucial to this research, as will the requirement for teacher input and feedback to ensure that the assessment is relevant for the purposes for which it is designed and that it is useful in the context it is intended to be used. As Baird et al. (2017) argued, accepting educational concerns when it comes to assessment does not need to result in rejecting concerns around assessment evaluation. Taking account of input and feedback from teachers who use the assessments will ultimately only increase the validity, usefulness and overall value of the assessments being used.

2.5 Conclusions

As this chapter has explored, there are many different types and methods of assessments relevant to use in special schools for varying purposes. In practice, the distinction between types and purposes of assessments may not always be obvious and some overlap may exist. Recent changes to statutory assessment in England mean that summative statutory assessment shifts focus to academic and pre-academic English and maths skills at a relatively early stage of functioning. These statutory assessments are used for pupils with numerous different intellectual and developmental disabilities and do not take account of the unique and specific needs of pupils on the autism spectrum with coexisting intellectual disabilities. For areas outside of these statutory assessments, however, there is scope for special schools to use assessments to identify and monitor progress which are best suited to the needs of their pupils. Pupils on the autism spectrum with coexisting intellectual disabilities are a distinct group with specific educational needs discussed in Chapter 1, such as in the areas of communication, RRBs, sensory needs and behaviour that challenges. These are also areas which may result in significant difficulties for this group of pupils to access and engage in learning opportunities in special schools and areas which teachers focus on throughout a pupil's schooling. Appropriate assessments are needed to identify skills, abilities and behaviours as well as monitor progress and support learning for pupils with autism and intellectual disabilities in the areas they often find most challenging. Due to the unique needs of these pupils, the assessments need to be autism-specific, robust, validated with appropriate samples and in relevant settings and useful to teachers.

Chapter 3. Description of the PhD

Within this chapter, the aims of this thesis and the different stages of the research will be outlined. The research gap within the literature will be clarified and the overall research question and specific research questions for each stage of the research will be identified. The assessment development procedures, important aspects of the assessment purpose and content, and justification of the research methods chosen for each stage will be considered. The associated ethical considerations for each stage will also be outlined. Finally, some broad limitations of qualitative and quantitative research and how these were resolved in the different research stages will be discussed.

3.1 The Research Problems

3.1.1 The Research Gap

Pupils on the autism spectrum with coexisting intellectual disabilities are a distinct group with specific needs. The interaction between autism and intellectual disabilities results in specific educational needs for this group of pupils, described in detail in Chapter 1, and these may result in barriers to these pupils accessing and engaging in learning opportunities. Basic skills in many of these areas such as communication and adaptive behaviour are required before further learning can take place. The presence of certain kinds of RRBs, sensory needs and behaviour that challenges may also limit a pupil's ability to learn. Interventions in schools may support pupils to learn to regulate extreme forms of these kinds of behaviours, not only to access learning, but also to improve outcomes and quality of life.

As explored in Chapter 2, statutory assessments in special schools are developed for pupils with a wide range of intellectual and developmental disabilities and do not specifically account for the specific needs of autistic pupils with coexisting intellectual disabilities. Teachers may need to identify abilities and competency in these areas before curriculum learning can take place and, if pupils require support to gain skills or regulate behaviours, then progress in these areas needs to be recorded and monitored. Schools have been given the flexibility to assess pupils in the way that is most appropriate for their needs and, therefore, there is an opportunity to use robust and useful autism-specific assessments which will allow teachers in special schools to

identify, monitor and record progress in barriers to learning for this group of pupils. The overall aim of the current research was to address this research gap by identifying which assessments, if any, are robust and can be used for this purpose, and to then undertake a series of studies leading to the development of an assessment specifically for use by teachers in special schools to assess autistic pupils with intellectual disabilities.

3.1.2 The Research Question

The main research question for this thesis is:

> How can the barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities be effectively identified and monitored through robust assessment by teachers in special schools?

This broader research question will be broken down into more specific research questions for each stage of the research. These research questions will be identified when each study is described.

3.2 Assessment Development

Although assessment development methodologies can be complex and may vary depending on a number of factors, they usually follow a number of basic steps including item generation, theoretical analysis and measurement property evaluation (Morgado et al., 2018). These steps involve the definition of the construct; clarification of the assessment method, purpose and relevant population; determination of the assessment model; development of content and items; and evaluation of measurement properties.

This research is divided into three stages. The first stage (Stage 1) of this research consisted of a systematic review to identify existing assessments in the peer reviewed literature relating to the areas of interest described in Chapter 1. The assessments were then evaluated for their appropriateness to assess outcomes for pupils on the autism spectrum with coexisting intellectual disabilities by teachers in special schools. In the second stage of this research (Stage 2), the assessment construct was developed and refined using the findings from the systematic review and professional knowledge of special needs teachers obtained through focus groups. Once the construct and assessment method were clarified, the second part of Stage 2 involved the

selection of items by special needs teachers using a Delphi exercise. In the final stage of the research (Stage 3), the assessment was piloted and results analysed to evaluate aspects of its reliability and validity. These stages will be described in further detail in the sections to follow.

3.2.1 Assessment Purpose

It is crucial to outline the assessment purposes prior to the development of the assessment as discussed in Chapter 2. The focus of this research will be the outcome or progress assessment of primary or secondary aged school pupils on the autism spectrum who attend a special school in the United Kingdom. As many autistic pupils are educated in mainstream settings, the population of interest in this thesis will have additional learning needs to the degree that they need educational provision outside of a mainstream school and are therefore likely to have moderate to profound intellectual disabilities. The assessment must be able to be completed, scored and interpreted by special needs teachers and must be appropriate for use in school, therefore focusing on skills, abilities and behaviours relevant to a school or classroom setting. The assessment must identify relevant skills and behaviours these pupils can show or that they are yet to develop, as well as monitor change in these areas. Decisions around the assessment method, construct and assessment model will be further explained in detail in the PhD outline below.

3.2.2 Assessment Content

Assessment development can follow a deductive process, an inductive approach or a mixture of the two. In a systematic review on scale development methodologies by Morgado et al. (2018), 56.2% of included studies reported using both deductive and inductive approaches. Deductive methods include use of the literature, established theory or existing measures to develop an assessment. Inductive methods may comprise of focus groups, interviews or observations in order to define the construct and generate items (Boateng et al., 2018). Boateng et al. (2018) suggested that this combination of the two methods is best practice in scale development and this research utilised both deductive and inductive methods in the development of the assessment content.

3.3 Epistemological Perspectives

A mixed methods research design was utilised within this project. Quantitative and qualitative methodologies reflect philosophical and epistemological research positions as well as the technical aspect of data collection (Rist, 1977). Johnson and Onwuegbuzie (2004) considered that rather than qualitative versus quantitative, described as the "incompatibility thesis", mixed methods research allows for a combination of qualitative and quantitative methodology which utilise the strengths and minimise the weaknesses of each design. Positivism or realism is the research paradigm associated with quantitative research methods which has an objective research focus with a view to find the truth or reality. Constructivist or interpretivist approaches, in contrast, aim to understand the subjective realities that individuals or groups construct and usually involve more qualitative methods and analysis (Brundrett and Rhodes, 2013). Proponents or purists for the positivist and constructivist paradigms, associated with quantitative and qualitative research methodologies respectively, each reject the other (Johnson and Onwuegbuzie, 2004). This can result in research methods being restricted by an attachment to one epistemological position, creating barriers to potentially effective research methods (Yvonne Feilzer, 2010). As Bryman (1984) argued, "there is no necessary 1:1 relationship between methodology and technique in the practice of social research" (p.89). Johnson and Onwuegbuzie (2004) suggested pragmatism as a way to bridge these paradigms and allow research questions to be answered in the best possible ways. They looked to the ideas of classical pragmatists Charles Sanders Pierce, William James and John Dewey to conclude that the empirical and practical consequences of ideas should be considered when judging and selecting research methods and making decisions about next steps. Their advocacy of the pragmatist perspective focused upon its practical *usefulness* in research and identified pragmatism as a "middle position philosophically and methodologically" (Johnson and Onwuegbuzie, 2004, p.17). Pragmatism is appropriate for the current research as there is a key focus throughout that assessments work in practice and are *useful* for teachers in the classroom.

The "very fact of educational research being multi-paradigmatic" results in the need for flexibility when answering different parts of a research question (Rist, 1977, p.42). In the current research, the development of the assessment content utilised both qualitative and quantitative research methods. The definition of the construct, 'barriers to learning', and the broader enquiry

into the appropriate areas, skills and behaviours for inclusion in the assessment, involved focus group discussions exploring teachers' opinions on what they believe to be important in assessment. This aligned more towards an interpretivist or constructivist position, acknowledging that each individual constructs their own reality (Bryman, 1984). Items for inclusion were determined by a Delphi study; a research method which "straddles the qualitative and quantitative divides" (Amos and Pearse, 2008, p.98). The Delphi exercise as a pragmatic research method allows groups of experts to reach a consensus, seeking the 'truth' yet accounting for the context of this 'truth' by acknowledging the experiences and subjectivity of the participants. Alongside this, a qualitative element was included in the Delphi exercise responses in line with a pragmatic position. Deeper and richer qualitative data allows for a greater understanding of reasoning behind responses and also reflects practical inquiry, which iteratively leads to further action and avoids "simplistic responses to very complex issues" (Yvonne Feilzer, 2010, p.11). The evaluation of the assessment in Stage 3 was positioned from a more positivist perspective, requiring quantitative statistical analysis to find the objective reality of measurement properties (Yvonne Feilzer, 2010). The pragmatist position, however, allows for flexibility of approach when balancing the need for robust measurement properties with the necessity of an assessment tool which is useful for the purposes it was developed (Fayers and Hand, 2002). Therefore, teacher feedback was considered an important aspect of the assessment evaluation stage to ensure that the assessment is not only robust but also useful for the purposes for which it has been developed.

3.4 PhD Outline

3.4.1 Stage 1

3.4.1.1 Study 1 – Systematic Review

The systematic review, along with the review of the literature in Chapter 1, made up the deductive aspect of the assessment development. A systematic literature review was first conducted to identify and evaluate available appropriate assessments and to ascertain any lack of available assessment tools used for the relevant purposes. Boateng et al. (2018) specified that a thorough review of the literature is necessary in assessment development to ensure that measures serving the desired purpose do not already exist. A systematic review may also identify gaps in

the literature and provide justification for further research (Moher et al., 2009). Best practice in the conduct of systematic reviews reduces bias when collecting and synthesising data and allows confidence in the evaluation of the evidence (Pussegoda et al., 2017). A systematic review was considered appropriate at this stage of the research to synthesise the information on the identified assessments and to present the information in an accessible way (Higgins et al. 2019). Although the systematic review in Stage 1 of this research was not related to interventions, it was nonetheless conducted in accordance with the Cochrane (Higgins et al., 2019) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) guidance. Prior to the systematic review being conducted, the PROSPERO database was searched in order to ensure that the review was not replicating previous research. The systematic review protocol was registered with PROSPERO (Prospero ID: CRD42019124827) and the record was updated as the review was conducted. The quality appraisal of the assessments included in the systematic review was guided by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) manual for systematic reviews of patient-reported outcome measures (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). The COSMIN checklist was created specifically for use in systematic reviews to evaluate the risk of bias in studies on measurement properties of patient reported outcome measures (PROMs) (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). Although the systematic review conducted in Stage 1 was concerned with outcome measures which are not specifically PROMs, the COSMIN checklist was used as a framework to assess both the methodological quality of studies and the reported properties of the assessments. The manual acknowledges that the COSMIN methodology may need to be adapted for uses outside of the specified areas (e.g. for clinician-rated measures) and further detail of the adaptations will be outlined in Chapter 4.

The two primary research questions for the systematic review conducted in Stage 1 of the research are as follows:

1. Which assessment tools can be used by teachers within special education settings to measure adaptive behaviour, behaviour that challenges or autism-related behaviour of children with intellectual disabilities?
2. Which of those assessment tools are appropriate for measuring the progress and

outcomes of children on the autism spectrum with coexisting intellectual disabilities within a special education setting?

As part of the evaluation of the appropriateness of identified assessments, a secondary aim was to evaluate the measurement properties of the assessments in order to judge their likely utility. The research questions for the systematic review related to children with intellectual disabilities for thoroughness, as measures appropriate for assessing those with intellectual disabilities may also be relevant to and are often used with autistic individuals with coexisting intellectual disabilities. The appropriateness of assessments for assessing skills and behaviours of autistic pupils was considered at the data extraction and discussion stages. Further details of the systematic review procedure are outlined in Chapter 4. This systematic review has been published (Howell et al., 2021; Appendix 1.1).

3.4.2 Stage 2

In Stage 2 of the research, the construct to be measured was defined and the assessment content determined. The inductive elements of the assessment development were introduced during Stage 2 and these included gathering special needs teachers' opinions through focus groups as well as selecting items to include using a teacher Delphi exercise.

3.4.2.1 Study 2a – Focus Groups

A number of research methods were considered in order to obtain teachers' opinions and further define the construct including interviews, questionnaires and focus groups. Focus groups were chosen for the first part of Stage 2 for a number of reasons. Focus groups allow participants to discuss, explore and clarify their ideas in real time as well as respond to others' contributions (Morgan, 1998). Some main features of focus groups include the face-to-face group interaction generating a "deeper and richer" range of data than one-to-one interviews along with the open-ended nature of the discussions compared to a questionnaire or survey (Rabiee, 2004, p.656). At the initial stage when generating ideas for the assessment, the interactive element allows participants to consider a variety of suggestions and opinions and enables them to explore and convey their priorities and areas of importance to them (Detmar et al. 2006). Focus groups can empower participants and it was thought that teachers would benefit from being recognised as 'experts' in their specialism and being given a voice in the area of assessment which is usually

prescribed at government or school level. The group structure of focus groups also helps to shift the balance of power and influence away from the researcher which may give space for participants to be more honest (Ivanoff and Hultberg, 2006). A final justification for using focus groups rather than one-to-one interviews was that they are less of a time burden on both researcher and participants. The focus groups were run at the schools and during staff meeting or directed staff time which would not have been possible with one-to-one interviews. It is likely that for one-to-one interviews, teachers would have been asked to participate outside of working hours and potentially even travel to do so. This was considered an unnecessary time burden which could not only reduce participation but also would create unnecessary pressure on teachers who already face a heavy and sometimes unmanageable workload (Walker et al., 2019).

The research aims of the focus groups were:

(a) to identify important areas of progress and barriers to learning for pupils on the autism spectrum to further refine the assessment construct,
(b) to consider broad areas and specific skills or behaviours which special needs teachers think are important to assess for autistic pupils and
(c) to consider the features of assessment tools which are useful to special needs teachers in the classroom.

The details of the focus group procedure are described in Chapter 5. This work has been published (Howell et al., 2020a; Appendix 1.2)

3.4.2.2 Construct Definition, Assessment Method and Assessment Model

There are many factors which may affect an individual's ability to learn, particularly for autistic pupils with intellectual disabilities. The construct defined in this research, 'barriers to learning' for pupils on the autism spectrum in special schools, was informed through the published literature but ultimately determined by the priorities of the teachers who teach these pupils. Following the focus groups in the second stage of this research, the construct was able to be more clearly defined. Barriers to learning was broadly defined as 'behaviours or skill gaps which can restrict or limit a child's access to or engagement in learning opportunities in the classroom or school environment'. More specifically, these areas were divided into behaviours or needs that were present and skills that were absent. Examples of additional needs or present

behaviours that the pupils may display include hypo- or hypersensitivities and RRBs. Examples of skill gaps due to atypical skill developments include the ability to regulate behaviour and emotions, difficulties with functional communication and skills that are prerequisites to learning such as attention, focus and engagement. These areas are not considered exhaustive and they may be broken down into smaller aspects of behaviour or more specific skills. The barriers may be consistently present (or absent) or they may be context dependent. Teachers identified that barriers to learning often include areas which need to be addressed before the focus can be on teaching further skills, especially curriculum teaching and learning. Improvements in these barriers to learning are likely to improve school functioning, future outcomes and both individual and family quality of life (Burgess and Gutstein, 2007).

Following the teacher focus groups, the assessment method chosen to measure this construct was a teacher rating scale and this was selected for a number of reasons. Firstly, report measures which require teacher respondents are in line with other assessments used in special schools and take account of the complex needs of the target population. Secondly, a rating scale would account for small changes and variation in skills in the way a dichotomous checklist could not. Also, the rating scale was designed to allow respondents to consider their wider knowledge of the pupil. This would mean that the assessment can be completed more quickly and flexibly than an observation or performance assessment as discussed in Chapter 2.

Finally, the list of items for the Delphi exercise was compiled based on the areas which teachers discussed in the teacher focus groups and this will be further described in Chapter 7. At this stage once the construct was defined, it was also necessary to identify the assessment model as being either reflective or formative (Jarvis et al., 2003; Coltman et al. 2008). The term ‘formative’ in the context of measurement models is distinct from the formative assessment purpose in educational assessment described in Chapter 2. For clarity, formative in this context will be specified as ‘formative model’. Diamantopoulos et al. (2008) defined a measurement model as something which “describes relationships between a construct and its measures” (p.1204). As discussed in Chapter 2, a construct is the unobservable or unquantifiable concept of interest while measures or items are the observed, quantified variables or indicators of the construct (Coltman et al. 2008; Edwards and Bagozzi 2000; Tay and Jebb, 2017). A reflective measurement model refers to a relationship which flows from the construct to the indicators and

certain items are therefore selected to reflect the construct. In a formative model, the relationship flows in reverse with the items combining to form or 'cause' the construct (Diamantopoulos et al., 2008; Coltman et al. 2008). Quality of life and socio-economic status measures are examples of formative assessment models as they are made up of factors (presented as items) which together 'create' these constructs (Diamantopoulos et al., 2008). Theoretical and philosophical debates on reflective vs formative models are outside of the scope of this research, however it is important to correctly identify the measurement model because the evaluation of measurement properties will be dependent upon whether the model is formative or reflective (Coltman et al., 2008; Streiner, 2003; Mokkink et al., 2016). Assumptions of unidimensionality, estimates of structural relationships and validity evaluation, for example, all depend on whether a measurement model is reflective or formative (Fleuren et al., 2018). Diamantopoulos et al. (2008) commented that misspecification of measurement models usually involves wrongly classifying formative models as reflective rather than the reverse. From a position of pragmatism, selecting the appropriate measurement model for the purposes outlined in this research is necessary for effective evaluation. Coltman et al. (2008) provided a framework for determining whether a measurement model is reflective or formative. The measurement model in relation to the barriers to learning construct was carefully considered and strong arguments were present for each. An assessment which measures barriers to learning was ultimately considered to be reflective, as a pupil's barriers exist independently of whether they can be observed at any given time and any change in the construct will result in a change in the observable indicators, i.e. skills and behaviours. Item correlation was a potential factor that pointed towards a formative model being appropriate for this construct, as it was thought that the items may not be interchangeable nor strongly correlate in a way which is required by a reflective model (Jarvis et al., 2003). However, this may be indicative of a multi-dimensional reflective model made up of a number of separate subscales. As reflective measurement models are preferable where possible due to the limitations of analysis of formative models, the assessment of the construct barriers to learning in this research was judged to be reflective (Fleuren et al., 2018).

3.4.2.3 Study 2b – Delphi Exercise

Following the focus groups, further definition of the assessment construct and the selection of the assessment methods and model, a list of items were created. A two round

modified Delphi exercise was then conducted in order to determine the items to be used in the assessment. The purpose of the Delphi exercise is for a group of participants to anonymously indicate agreement via questionnaire, receive feedback on group responses and then refine their answers in successive rounds, driving towards a consensus on the issues in question (Dalkey and Helmer, 1963).

The Delphi exercise is an established method of evaluating content validity of assessments by determining consensus of experts on the items and content (Boateng et al., 2018). In contrast to the focus group, a feature of the Delphi methodology is that the participants do not have any face-to-face interaction (Green, 2014). The Delphi method was chosen for the item selection stage of the research as the focused and quantitative data which are produced for each assessment item allowed for clear decision-making about the rejection or retention of items. Although quantitative analysis is the main data analysis method in a Delphi exercise, a comments section was also provided for each item if participants wished to provide further justification for their responses or suggestions for modification. The Delphi is suitable for this research as it is a method which is useful in bridging the gap between research and practice (Stone Fish and Busby, 2005). The Delphi method was also considered convenient for teacher participants as they could contribute anonymously, did not have to be physically present and could complete the questionnaires at a time and place of their choosing (Green, 2014).

The aims of the Delphi exercise were to consider:

(a) the relevance of a list of potential assessment items,
(b) the comprehensibility of those items and
(c) the comprehensiveness of the list of items by addressing whether any skills or behaviours were missing from the list.

The original Delphi method has been adapted and modified for use in a range of disciplines and covering a wide variety of research designs and aims (Hasson and Keeney, 2011). The Delphi procedure was modified for the purposes of this research and the modifications will be further outlined in Chapter 6. This study has been published (Howell et al., 2020b; Appendix 1.3)

3.4.3 Stage 3

3.4.3.1 Assessment of Barriers to Learning in Education – Autism (ABLE-Autism)

Once the list of items had been finalised through use of the Delphi method, the assessment was created. The assessment was named the Assessment of Barriers to Learning in Education – Autism (ABLE-Autism). The assessment abbreviation ABLE-Autism was selected for positively reflecting the research focus on supporting pupils on the autism spectrum to make progress and overcome barriers. The structure and features of the assessment will be detailed in Chapter 7.

3.4.3.2 Study 3 – Assessment Evaluation

Methods used to evaluate an assessment may vary according to the information required. Relevant aspects of reliability and validity may depend upon the assessment purposes and population. Evaluations may also be dependent upon recruiting a large enough sample to conduct the analysis (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). For the current assessment evaluation, test-retest reliability, inter-rater reliability, internal consistency reliability, convergent validity and responsiveness were considered to be the most appropriate measurement properties to evaluate. Factor analysis could not be conducted due to the requirement of a sample size at least five times the number of items, which would require 350 participants (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). IRR was considered an important measurement property to evaluate, however there were a number of practical difficulties evaluating this aspect of reliability. Similarly, responsiveness was considered an important measurement property to evaluate for an outcome assessment, however the responsiveness of the assessment could not be determined due to U.K. schools closing as a result of the COVID-19 pandemic. The difficulties with IRR and responsiveness evaluations will be discussed further in Chapter 8.

Many scale validation procedures consider that the data are produced through independent dyads (i.e. one teacher assessing one pupil). However, in education it is common for one teacher to assess a number of pupils. Multilevel modelling was therefore used to account for the nesting of data and the fact that not all pupil-teacher dyads were independent. In the multilevel model, pupils were the unit of analysis (level one) and teachers were the nesting variable (level two). The inclusion of an additional level accounting for pupil data nested within

teachers nested within schools was considered. However, as fewer than half the participating schools had assessments completed by more than one teacher, the model which accounted for teacher effects was considered appropriate. Internal consistency reliability, test-retest reliability and convergent validity with a similar measure (the Teacher Autism Progress Scale; TAPS) were calculated accounting for the nesting of data. The individual procedures will be described further in Chapter 8. This work has been submitted for publication and is currently under review (Appendix 1.4).

3.4.4 Limitations and Risk of Bias

3.4.4.1 Limitations of Qualitative Research

Limitations of each research method will be discussed in detail in the study chapters to follow. However, there are some more general limitations to the different research approaches.

Qualitative research has been criticised for lacking transparency and scientific rigour and for the influence of subjective bias by the researcher (Noble and Smith, 2015). It may be argued that all qualitative methods may have an element of unavoidable bias, as the nature of interaction between individuals and the level of interpretation necessary in these exchanges is unlikely to ever be considered bias-free (Silverman, 2016). Morse (1994) argued that applying quantitative standards of validity and reliability to qualitative research "violate[s] the philosophy, purpose and intent or the qualitative paradigm" which is focused upon an in depth understanding of the qualities of the area under study (p.97). There is still debate as to whether and how validity and reliability should be applied to qualitative research (Hannes, 2011). The Cochrane Guidance for critically appraising qualitative research uses the following terms: credibility, whether the data truly represents the participants' views; transferability, whether the results are transferable; dependability, whether the research process and decisions made are justified and well documented; and confirmability, how the results can be confirmed through the data (Hannes, 2011).

In qualitative research, the researcher is an instrument in the data collection process (Ortlipp, 2008). As a former teacher conducting research with teachers, it is acknowledged that the researcher may be subject to unintended biases. One way to seek to address the possibility of any subjective bias of the research is through the use of a reflective journal. Including subjective

accounts as part of the research process may appear counter-intuitive to the aim of objectivity. Acknowledgement by the researcher of their part in the research process can, however, support the dependability and truthfulness of the research by providing justification for decisions made during the process and create an audit trail of the research decisions (Jasper, 2005). By acknowledging their history, experiences and preconceptions and making them visible, the researcher can guard against unidentified bias and ensure that the research process is as transparent as possible (Ortlipp, 2008). Reflective research practice also aligns with the pragmatism paradigm; questioning the research and the researcher's role and influence in the research process in order to ensure the outcomes are useful (Yvonne Feilzer, 2010). A reflective journal was used to support the credibility, dependability and confirmability of the research through the decisions, processes and data analysis.

3.4.4.2 Limitations of Quantitative Research

Quantitative research is the dominant method in assessment and measurement research. One of the main limitations of quantitative research is the lack of complex meaning which can be discerned from data; quantitative methods may show 'what' but often cannot answer the 'why' and 'how'. As explored in Chapter 2, measurement involves assigning numbers to concepts, traits and attributes which may not be directly measurable. Although the processes involved in quantitative data analysis seek to ensure that this measurement is as valid and reliable as possible, the quantitative results may not be enough to ensure that the assessment is valuable and useful to those who use it. For example, the quantitative data analysis might suggest the measurement properties of the assessment are sound, however teachers may consider a crucial aspect of the assessment content to be missing. Similarly, if the data suggest that aspects of the assessment, e.g. reliability, are not sound, it may not be clear how to address that without input form the teacher respondents. These limitations have been addressed through the use of mixed methods research (Hammond, 2005). The qualitative aspects of this assessment development and evaluation process have been designed to counter these limitations and to ensure that, at each stage, there are opportunities for teachers to comment and explain if necessary. Comments boxes have been included in both the Delphi study and the feedback associated with the assessment evaluation to ensure that the quantitative data can be supplemented by rich, qualitative data which supports the decisions made and ensures that the deeper meaning behind the numbers is

not lost. Referring to qualitative data helps to triangulate and validate quantitative findings (Onwuegbuzie and Leech, 2005). This also aligns with the pragmatist position of this research, which allows the complementary aspects of quantitative and qualitative methods to reduce the limitations of each.

3.5 Ethics

3.5.1 Ethical Approval

Ethical approval was sought from the Tizard Ethics Committee for each relevant stage of research prior to the research being conducted. The research was approved as set out below with relevant documentation provided in Appendix 3.

Stage 1 – No ethical approval required

Stage 2 – Approved 18/04/19 (minor amendments approved 03/06/19)

Stage 3 – Approved 03/10/19 (minor amendments approved 07/11/2019, 20/11/2019 and 22/01/2020)

3.5.2 Ethical Issues: Stage 2

There were a number of ethical considerations which were important to address prior to the commencement of the focus groups and Delphi exercise in Stage 2 of the research. These will be considered below in relation to the four categories set out in the framework of Principles of Biomedical Ethics (Beauchamp, 2007). Further detail of the procedures conducted in order to apply with these ethical considerations are outlined in the individual study chapters. Ethical amendments were sought after the first focus group to include a third focus group involving parents. However, due to time limitations, the parent focus group was not conducted and parent consultation on relevant areas and the assessment content will instead be a recommendation for future research.

3.5.2.1 Nonmaleficence (Causing No Harm)

The subject matter of the discussions and questionnaires was not of a sensitive nature and it was not anticipated that any harm, such as distress or embarrassment would be caused through participation. However, it was important that teachers were informed that their responses would

be confidential and that they could answer openly and honestly. The only disadvantages of participating included the time burden on participants and this was minimised as far as possible by conducting focus groups at the participants' place of work and during school or staff meeting time.

3.5.2.2 Beneficence (Doing Good)

It was considered that potential benefits of participation may include participants feeling heard and having their opinions and experience valued. Indirectly, teachers may have benefitted from the knowledge that they had contributed to a new assessment and that this assessment may be useful for supporting their classroom assessment practice in the future.

3.5.2.3 Respect

The main ethical considerations for the studies conducted in Stage 2 included respect for participants' autonomy and involved issues of consent, confidentiality and voluntary participation. It was necessary to consider that teachers may be under pressure to participate in the focus groups by their school managers. Therefore, voluntary participation and the unconditional right to withdraw was made explicit, not only in the participation information sheet and consent form but also verbally by the focus group moderator. Additionally, as some of the participants were part of the researchers' professional networks and therefore known to the researcher, care was taken to avoid social desirability bias, e.g. opportunities to please or agree with the researcher. Measures taken included explicitly informing the participants that no answers were right, wrong or expected and the fact that the researcher guarded against leading the discussion or agreeing with particular responses. Similarly, in the Delphi exercise, participants were informed that all responses were anonymous and that no particular responses were expected.

3.5.2.4 Justice

It was anticipated that the development of a new assessment as a result of this research will be to the advantage of teachers, pupils and parents. A useful assessment which acknowledges progress will allow schools and parents to celebrate achievements of their pupils. Teachers' workload may be reduced by having an effective assessment to show further progress in areas other than those covered by the curriculum. Schools will benefit by having further data

to support the effective teaching and learning. In addition, a further advantage of this research is the recognition of teachers as experts in their field and the opportunity for teachers to share their views of areas of practice that are usually prescribed, e.g. assessment. Additionally, the research may be advantageous in highlighting effective research methods which can be used to include teachers in the development of new assessments.

3.5.3 Ethical Issues: Stage 3

A number of ethical issues required consideration when designing the procedure for Stage 3 of the research, particularly due to the involvement of special school pupils. Many of the ethical considerations for Stage 2 were mirrored in Stage 3. This study required additional consideration of the ethical issues related to the inclusion of children in the research. Other minor ethics amendments approved on 07/11/2019 and 22/01/2020 related to the use of a specified assessment for convergent validity, recruitment of participants via social media and the adaptation of an email inviting headteachers to participate into a physical letter.

3.5.3.1 Nonmaleficence (Causing No Harm)

It was necessary to further consider the time burden for teachers participating in Stage 3, especially as the assessment contained a larger number of items following endorsement in the Delphi study than was initially expected. To reduce the time burden on teachers as much as possible when participating in the evaluation study, the additional assessment selected for the convergent validity analysis was intentionally short and the number of completions required was the lowest number required for analysis. Additionally, teacher participants were advised that there was no minimum number of pupils that they needed to assess to participate and that they could choose to assess additional pupils if they wanted to. The unconditional right to withdraw was also made clear on a number of occasions throughout the research and teachers were able to complete just the first round of assessments if they did not wish to continue participating further.

The children were not *directly* participating in the research as the assessment was a report measure using the teachers' knowledge of the pupils. Therefore, the children were not asked or required to do anything for the assessment and no risk of harm or burden to the pupils or parents was anticipated.

3.5.3.2 Beneficence (Doing Good)

The potential benefits for teachers, parents and pupils as a result of participation in this research are likely to be indirect and include knowing that they have contributed towards the development of a useful assessment to support the identification and monitoring of pupils' barriers to learning. Further longer-term benefits of participating may include the future availability of a robust and beneficial assessment tool, developed through this research, which allows teachers to show progress that their pupils are making in the skills and behaviours which may affect their learning. With the future availability of the completed assessment, both teachers and parents are likely to benefit from being able to see that their pupils and children are progressing in these important areas and the opportunity to celebrate this will be a positive experience.

3.5.3.3 Respect

Although the children were not direct participants, information about the pupils' skills and behaviour was to be shared. It was therefore necessary for parents, and pupils if appropriate, to be informed of the research and to be given a chance to withhold consent at any time. To ensure confidentiality it was decided that, in line with the principle of data minimisation within the General Data Protection Regulations (GDPR), the pupil information would be anonymised by the teacher so that no pupil identifying information would be transferred to the researcher.

3.5.3.4 Justice

Advantages of Stage 3 of the research were the same as discussed above for Stage 2; a new, robust, autism-specific assessment that can be used by teachers in special schools to identify and show progress in pupils' barriers to learning will be beneficial for teachers, parents, pupils and schools.

3.6 Conclusions

Within this chapter, the structure of the PhD research was described, as were key arguments and the research question and aims. The processes used to develop the assessment measure were explained, including the principles and purpose of the assessment measure and content. The research methods to be used have been outlined and justified both in terms of the

philosophical position of the research as well as practical considerations. Ethical considerations were discussed and the procedures which ensured compliance with ethical research practices will be detailed in the following study chapters. This chapter considered some alternative research methods not chosen as well as reflecting on some of the broad limitations of qualitative and quantitative research and how mixed methods research attempts to reduce some of these limitations. Many of these issues will be expanded upon in later chapters. The following chapters will describe the individual studies leading to the development of the ABLE-Autism in detail.

Chapter 4. Study 1 – Systematic Review

4.1 Introduction

Following the review of the literature described in Chapter 1, a systematic review was conducted at the first stage of this research in order to identify whether any current assessments were appropriate for the research purposes. As explained in Chapter 3, this research method is useful to reduce bias and subjectivity during the systematic evaluation of the literature and the identification of research gaps.

4.1.1 Previous Reviews

Much of the research focus of assessment in the area of special needs and autism has been around diagnosis and screening rather than outcomes and progress (Wigelsworth et al., 2015). Three previous systematic reviews have been conducted on outcome assessments for autistic individuals. McConachie et al. (2015) conducted the first comprehensive review of the quality and appropriateness of progress and outcome measures for children on the autism spectrum. The review identified 131 assessment tools in total which were appropriate for this purpose. They found 128 papers which commented on the measurement properties of 57 tools in a number of different domains including autism symptom severity, social awareness, RRBs, sensory processing, cognitive ability, language, attention, behaviour problems and daily living skills. Strong evidence was found for 12 assessments, the majority of which assessed autism characteristics and 'problem' behaviour. The review aimed to provide information and guidance for practitioners to select assessment tools for their specific purpose rather than recommending or ranking the assessments (McConachie et al., 2015). A number of notable conclusions were made. Firstly, they identified the lack of literature reporting on the strengths and weaknesses of outcome measures for individuals on the autism spectrum. Secondly, it was noted that assessments which were originally developed for screening and diagnostic purposes were often used as outcome measures. Also recognised was evident "tension between the diagnostic process in ASD, and the focus on parent and professional valued outcomes" (p.xxviii). Additionally, they commented that norm-referenced assessments which compare autistic children to typically developing children may be misleading due to atypical developmental profiles and potential lack of cooperation. Finally, they found that few studies considered the responsiveness of the

assessment tools and also found little evidence that the outcome measures could detect change due to interventions (McConachie et al., 2015). These observations support the ideas discussed in Chapters 1 and 2; that autistic pupils are likely to have differing needs to children with other conditions when it comes to assessment and that current assessment practices may not always adequately reflect the needs of this group of individuals.

As an extension of the McConachie et al. (2015) review, Hanratty et al. (2015) conducted a systematic review looking at the measurement properties of assessment tools used to assess behaviour difficulties in young children with ASD under the age of six years. Fifteen studies were found which evaluated six assessment tools. The measurement properties of the Child Behavior Checklist and the Home Situations Questionnaire were found to be the most robust. In line with the review by McConachie et al. (2015), they found that evidence for the measurement properties of tools was patchy and also noted that responsiveness is a property which is often not considered in studies evaluating assessment tools even though it is a particularly relevant when measuring progress.

A recent systematic review by Provenzani et al. (2020) identified assessments used to measure autism-related outcomes in clinical trials. They found 327 outcome measures, 69% of which were only used within the literature once. Only seven assessments were used in over 5% of the studies. As was reported by McConachie et al. (2015), Provenzani et al. (2020) outlined the regular use of non-specific assessments for autism and noted that many of the assessments were not developed as outcome measures.

To the best of this author's knowledge, no published research has been conducted on the use of assessments by teachers in special education settings for autistic pupils with coexisting intellectual disabilities.

4.1.2 Research Aims

The specific educational needs of autistic pupils in special schools which were discussed in Chapter 1 were the focus of the initial barriers to learning construct due to their potential for limiting access to and engagement in learning. For the purposes of the systematic review, the initial areas of interest were condensed to form three categories; adaptive behaviour, behaviour that challenges and autism-related behaviour. The areas of RRBs, sensory needs and the uneven

cognitive profile were incorporated under the category of autism-related behaviour. The term ‘autism-related behaviour’ was preferred to language such as ‘autism symptomatology’ but included areas covered by this terminology. Communication was not included as its own category in order to avoid assessments related to the technical aspects of communication and language which go beyond the expertise of classroom teachers and require support from other professionals such as speech and language therapists. Functional communication skills were instead considered within the category of adaptive behaviour as many adaptive behaviour assessments contain items or subscales related to communication and social skills.

While McConachie et al. (2015) and Hanratty et al. (2015) identified a gap in the research on autism assessment, this systematic review differed in a number of ways. Firstly, this review is narrower in scope than the reviews by McConachie et al. (2015) and Provenzani et al. (2020) but broader than Hanratty et al. (2015). Secondly, this review extended the age range of previous reviews by including assessments appropriate for school aged children. Thirdly, the focus of this review differed from previous reviews by considering assessment tools and their measurement properties in the context of their applicability to a special education setting and their implementation by teachers. Additionally, this review included assessments devised for individuals with intellectual disabilities, such as those used in special schools, and considered them in relation to individuals on the autism spectrum with coexisting intellectual disabilities. Finally, although McConachie et al. (2015) and Hanratty et al. (2015) applied a date range of 1992 to 2015, this review will not limit inclusion by date and, instead, the date of the research will be a consideration in the analysis of the included studies.

4.1.3 Research Questions

This review addressed two primary questions:

1. Which assessment tools can be used by teachers within special education settings to measure adaptive behaviour, behaviour that challenges or autism-related behaviour of children with intellectual disabilities?

2. Which of those assessment tools are appropriate for measuring the progress and outcomes of children on the autism spectrum with coexisting intellectual disabilities within a special education setting?

As part of the evaluation of the appropriateness of identified assessments, a secondary aim was to evaluate the measurement properties of the assessments in order to judge their likely utility.

4.2 Methods

A search was conducted for studies which reported primary data on the measurement properties of assessments used to measure adaptive behaviour, behaviour that challenges or autism-related behaviour. As described in Chapter 3, the exact definition of what constitutes 'adaptive behaviour' is unclear and the behaviours measured by different assessments may vary (Kramer et al., 2012). For the purposes of this research, adaptive behaviour assessments are those which focused on assessing functional, applied or generalised skills including independence. To address the fact that there is no clear distinction between measures of 'participation' and adaptive behaviour, measures of participation were included if they were appropriate for a school setting, could be used by teachers and the focus was on skills or abilities relevant to participation as opposed to measuring levels of participation. To ensure all relevant tools were identified, all assessments appropriate for assessing individuals with intellectual disabilities were considered. These assessments were then evaluated in light of their application to pupils on the autism spectrum.

4.2.1 Searches

Searches were conducted of a number of electronic databases using EBSCOhost including Academic Search Complete; British Education Index; ERIC; MEDLINE; PsychArticles; PsychInfo and CINAHL. A separate search was conducted using PubMed. Table 1 shows the key search terms used including combinations, spelling variations and truncation.

Table 1. *Search Terms*

	Search term blocks combined with AND						
Search terms (OR)	AB(autis* OR "pervasive development* disorder*" OR	AB(assess OR assessment)	tool OR instrument OR scale OR	progress OR progression OR attainment OR	measure OR evaluate OR test	development OR functioning OR social OR	NOT(elderly OR dementia OR concussion OR

"learning disab*" OR "learning difficult*" OR "intellectual* impair*" OR "intellectual* disab*" OR "mental* disab*" OR "abnormal development" OR "developmental* disab*" OR retard* OR handicap* OR "special needs" OR "special educational needs")	checklist OR questionnaire	skill* OR outcome*	emotional OR communication OR behav* OR motor OR sensory OR math* OR numeracy OR English OR literacy OR language OR interaction OR play	dizziness OR dyslexia OR midwifery OR "drug therapy" OR "health related quality of life" OR "HRQOL" OR "life satisfaction" OR cancer)

The search yielded 3,497 results. The search was then repeated with PubMed and 323 articles were found. Automatic removal of duplicates resulted in 2,397 articles and a hand removal of duplicates left 2,270 articles for consideration.

Different combinations of the above search terms were also used to search the grey literature using opengrey.eu but no relevant results were found. The above databases were also used to search for assessments commonly used in special schools by name (e.g. Early Years Foundation Stage Profile, P Scales, B Squared) but, again, no relevant results were found.

4.2.2 Eligibility Criteria

Articles were first screened by title and abstract according to the inclusion and exclusion criteria in Table 2. Eligibility was not restricted by year of publication.

Table 2. *Eligibility Criteria*

Inclusion Criteria	Exclusion Criteria
1. The focus of the study was to evaluate one or more properties of an assessment tool (either a newly developed or existing assessment tool). 2. The assessment was intended for use with and was evaluated using a sample of which some individuals had intellectual disabilities (IQ < 70), or both autism and intellectual disabilities. Where general intellectual functioning of the sample could not be determined, the authors were contacted for clarification. Where no response was received, if the study specified that the participants attended a specialist school/education setting or had educational needs potentially unsuitable for a mainstream school (e.g. nonverbal/limited language) it was included. 3. The assessment was evaluated using a sample which overlapped with school age range (≥4 years and ≤16 years).	1. The study described the sample as having Asperger's Syndrome, 'high-functioning autism', specific learning disabilities (e.g. dyslexia) without coexisting intellectual disabilities or specified that all participants had an IQ > 70. If a study reported that the average IQ of participants was > 70 (without a range) or if it was not possible to determine whether the participants had intellectual disabilities, the study was excluded. 2. The assessment was evaluated primarily for its properties relating to diagnosis, screening for likely diagnosis, its ability to identify individuals with intellectual disabilities/autism or predict a future diagnosis. 3. The assessment tool was developed to be part of or used to assess a specific curriculum or covered the areas of interest but focused on a broad range of developmental outcomes, limiting its generalisability for use in other contexts. 4. The assessment was devised to be used specifically by health professionals. 5. The focus of the assessment is on the participation of the individual rather than their skills, abilities, performance or behaviour.

4. The focus of the assessment was to measure outcomes related to adaptive behaviour, behaviour that challenges or autism-related behaviour. 5. The study was a primary quantitative or qualitative study with 5 or more participants. 6. The study was a full text study published in a peer-reviewed journal and published in English.	6. The assessment is subject to publisher qualification codes for purchase (e.g. by a qualified psychologist) and is not freely available or able to be purchased, administered and scored and the results interpreted by a qualified teacher. 7. The assessment tool was used to measure outcomes for offenders with intellectual disabilities. 8. The assessment was used to measure outcomes for a sample with mental health conditions (e.g. depression, anxiety, psychosis, Attention Deficit Hyperactivity Disorder, Reactive Attachment Disorder, anger) or, if a sample with coexisting intellectual disabilities and mental health conditions was used, the mental health condition was the focus of the study. 9. The assessment was evaluated using a sample specifically with traumatic brain injury. 10. The assessment was used to assess the quality of life, the function of behaviour that challenges or to assess risk. 11. The assessment tool was developed as an outcome measure for a particular study and was based on observation of characteristics specific to that study (e.g. on task behaviour, parent-child interaction, coded observation) and could not be searched for by name. 12. The assessment tool was predominantly employment, vocation or training based. 13. The focus of the study was predominantly to evaluate a translated tool. 14. The article reported a case study, conference abstract or test review.

4.2.3 Screening

Following the first title and abstract screening, 2,196 studies were excluded leaving 74 studies included for a second screening.

The second screening determined whether the assessment met criterion 6 of the exclusion criteria. Criterion 6 excluded assessments which are subject to publisher qualification codes for purchase (e.g. by a qualified psychologist) and are not freely available or able to be purchased, administered and scored and the results interpreted by a qualified teacher. Twenty-eight studies were excluded at this stage because the assessment had a publisher qualification code which required a clinical psychology qualification in order to purchase or use (e.g. Pearson Clinical codes CL1, CL2; WPS publishing Level N) and therefore could not be used by a teacher. Some education systems (e.g. some U.S. states) require a master's degree or further training in SEN in order to teach this population. However, this is not a requirement for special needs teachers in England who can teach in special schools with an ordinary teaching qualification. A number of U.S.-based publishers have intermediary qualification codes which reflect this requirement of further qualification in order to purchase and use specific assessments; these measures were included and their utility will be considered within the discussion.

The full text article was obtained for 46 studies.

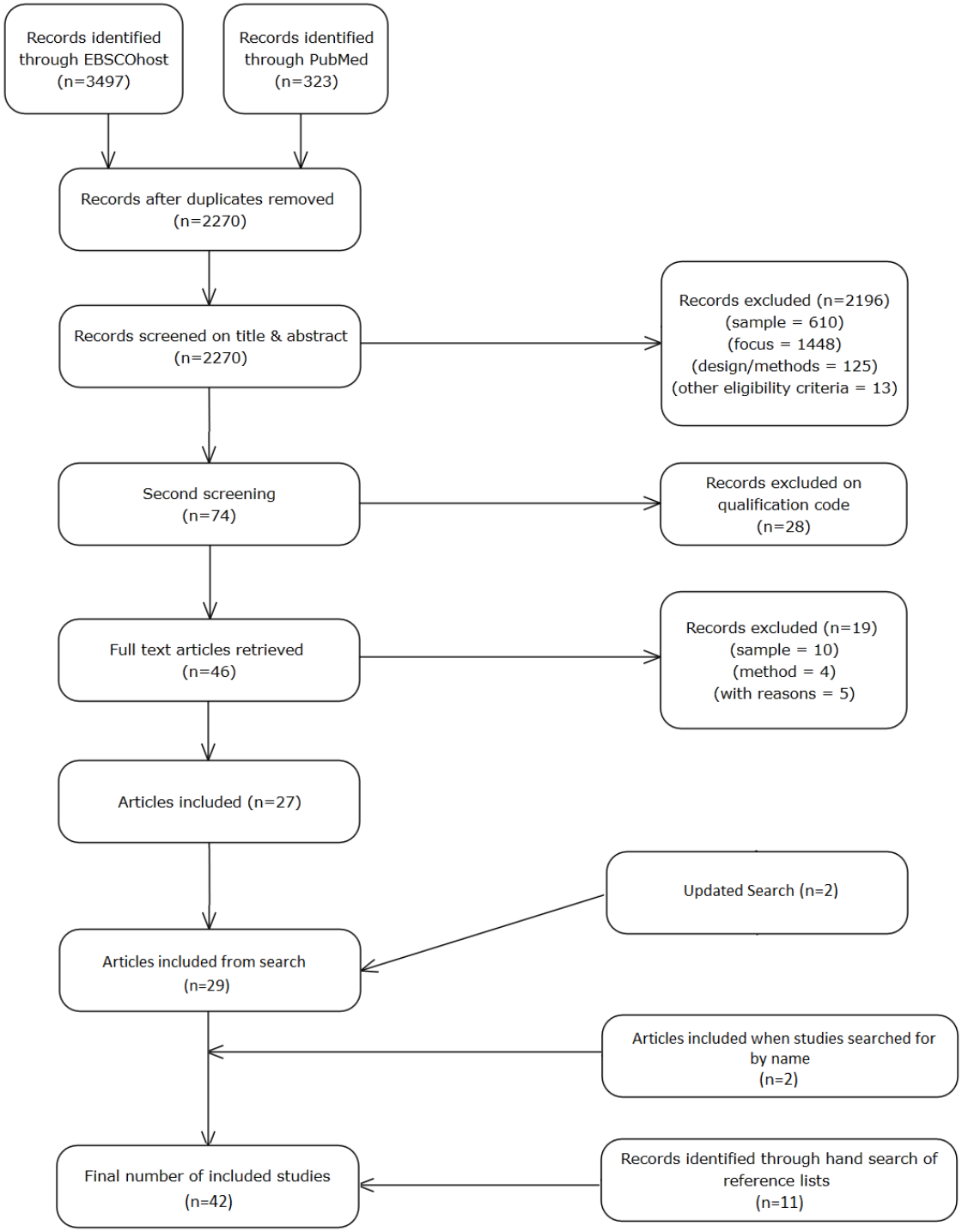
On full text screening, 19 studies were excluded. Seven were excluded due to sample age or absence of intellectual disabilities and four were excluded due to ineligible methodology. A number of authors were contacted for further information about the study. Three studies were excluded because IQ or intellectual disabilities of the participants could not be discerned, there was no suggestion that the sample required any special education provision and there with no response from the study authors. Five studies were excluded for not measuring relevant specific domains. The total number of studies included from the search was 27. The search was updated twice in 2019, and two additional studies matching the inclusion criteria were found, bringing the number of included studies from the search to 29.

The author became aware of an assessment from the wider literature which was particularly relevant for the purposes of this review but was not found through the search. This assessment, the School Function Assessment (SFA), was searched for by name using the same

databases and combined with the first block of search terms shown in Table 1 (variations of ‘intellectual disabilities’). Two studies were included as per the inclusion criteria.

The ancestry method identified a further 11 studies eligible for inclusion. Although including a large number of extra studies through a manual search may signify limiting search terms, eight of these studies commented on assessments already included through the search. In total, 42 articles reporting on 26 assessment tools were included. The search is outlined in Figure 1.

Figure 1. *Prisma Flow Diagram.*



4.2.4 Risk of Bias and Study Quality

As mentioned in Chapter 3, the quality appraisal was guided by the COSMIN manual for systematic reviews of patient-reported outcome measures (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). The COSMIN checklist covers aspects of development, validity, reliability and responsiveness with the methodological quality rated separately for each measurement property as 'very good', 'adequate', 'doubtful' or 'inadequate'. The lowest rating in each box is taken to be the overall rating of the quality of evidence for each measurement property. The COSMIN manual also contains criteria for sufficient measurement properties.

While much of the COSMIN checklist is highly relevant to the appraisal of tools for use in an educational setting, this review was concerned with the appropriateness of outcome measures in schools rather than clinical practice. As suggested by the COSMIN manual, the checklist was adapted for the purposes of this study and therefore some qualifications to the checklist must be made. Doubtful or inadequate ratings were, in some cases, due to missing information where studies did not provide sufficient detail required by the COSMIN checklist for a high rating. As different categories of assessment were considered and many assessments measure slightly different aspects of behaviour, it was more informative for these purposes to consider convergent validity in relation to each hypothesis rather than the criteria specified for criterion validity. Correlations with other measures for comparison were therefore appraised with reference to convergent validity. In addition, the COSMIN standards of similar test conditions for the purposes of reliability did not account for different versions of assessments with different respondents (e.g. teacher and parent forms) and therefore teacher-parent reliability was evaluated with reference to convergent validity as opposed to IRR. Similarly, studies on different modes of administration (e.g. telephone administration by interview with written reports) were considered relevant and coded with reference to convergent validity. Minimally important change (MIC) is not considered an adequate measure of responsiveness according to the COSMIN checklist so in order to consider MIC for the current purposes as reported in Chatham et al. (2018) this was rated with reference to hypothesis testing. As the purpose of this review was to report on outcome measures, correlations with IQ and diagnostic tools were not considered in light of the evaluations of measurement properties. Although the COSMIN manual suggested addressing scores individually for each subscale, this review reported on assessments' properties overall

where possible with comments on subscale results as necessary. The COSMIN manual allows for additional criteria to be used when assessing results from exploratory factor analysis (EFA) and this review used the criteria for sufficient construct validity outlined in McConachie et al. (2015); that factors explain > 50% of the variance. The COSMIN manual outlines a way of pooling or summarising results per measurement property, per assessment. Even where assessments had more than one study reporting on their properties, only a small number of studies considered the same measurement properties and these often utilised different versions of the assessment. As further information is unlikely to be provided by summarising the quality of the assessment as a whole, the results, therefore were reported and discussed separately for each study as well as an overview provided in table format. When discussing the implications of the quality assessment, while much of the COSMIN checklist is highly relevant to the appraisal of tools for use in an educational setting, it is necessary to consider that this review is concerned with the appropriateness of outcome measures to educational settings rather than clinical practice. Properties such as predictive validity and cross-cultural validity were not considered for the purposes of this review. Therefore, the COSMIN checklist was used as a guide and the quality ratings of studies, the evaluation of measurement properties of the assessments and the appropriateness of their use by teachers in schools were considered and discussed in the results section of this chapter. Twenty-one percent of the studies were rated by a second blind rater. The ratings were then reviewed and errors were corrected. The inter-rater agreement was 94%, and $k = .85$.

4.3 Results

4.3.1 Description of Included Studies

In total, 26 assessments were evaluated by 42 included studies and participant numbers ranged from 14 to 9067. Where studies evaluated different versions of an assessment (e.g. parent and teacher versions) or used separate samples (e.g. typically developing sample and sample with intellectual disabilities, adult sample and child sample) only data relevant to the eligibility criteria were considered unless comparisons were relevant (e.g. comparisons between parent and teacher responses). The data extracted from the studies are combined in Table 3.

Table 3. *Data from Included Studies*

Assessments of Adaptive Behaviour and Behaviour that Challenges											
Assessment	**Study/ Year**	**Country**	**Category**	**Sample**	**Respondents**	**Assessment Type**	**Newest/ Updated version?**	**Available?**	**Appropriate for Teachers?**	**Appropriate for School?**	**Intended population and purpose**
AAMD Adaptive Behaviour Scale *- Public School Version*	Mayfield et al. 1984	USA	Adaptive behaviour and behaviour that challenges	*n* = 31 Intellectual disabilities IQ 55-70 7-12 years old	Teachers (SEN and regular), parents and independent researcher	Interview or questionnaire	AAMR ABS- II	Unknown	Yes	School version evaluated with teachers	Intellectual/ developmental disabilities and autism Screening and evaluating skills
AAMD Adaptive Behaviour Scale *- School Edition*	Perry and Factor 1989	Canada	Adaptive behaviour and behaviour that challenges	*n* = 15 Autism IQ 18-82 8-18 years old	Parents and group home workers	Interview or questionnaire	AAMR ABS- II	Unknown	Yes	School Edition evaluated	Intellectual/ developmental disabilities and autism Screening and evaluating skills
AAMD Adaptive Behaviour Scale *- Part 2*	Spreat 1982	USA	Adaptive behaviour and behaviour that challenges	*n* = 221 Intellectual disabilities IQ 5-76 9-57 years old	Direct care technicians (staff)	Interview or questionnaire	AAMR ABS- II	Unknown	Yes	No mention of teachers/use in schools	Intellectual/ developmental disabilities and autism Screening and evaluating skills
AAMR Adaptive Behavior Scales *- School 2nd Edition (revision of AAMD school edition)*	Wells et al. 2009	Canada	Adaptive behaviour and behaviour that challenges	*n* = 50 Autism and intellectual disabilities 5-37 years old	Staff	Questionnaire or structured interview	Unknown	Abledata out of print 2016	Yes	School version evaluated	Intellectual/ developmental disabilities and autism Screening and evaluating skills

Behavior Assessment System for Children - 2	Ellison et al. 2016	USA	Adaptive functioning and internalising and externalising behaviours	*n* = 67 IQ 38-128 26 months to 18 years old	Teachers and parents	Rating scale	3rd Edition	Pearson Clinical	Level B	Teacher rating scales evaluated	Developmental/intellectual disabilities Screening/ diagnosis and evaluating/ measuring behaviour
Behavior Assessment System for Children -2	Lane et al. 2013	Aus	Adaptive functioning and internalising and externalising behaviours	*n* = 39 Autism in early intervention 2-5 years old	Teachers and parents	Rating scale	3rd Edition	Pearson Clinical	Level B	Teacher and parent version evaluated	Developmental/intellectual disabilities Screening/ diagnosis and evaluating/ measuring behaviour
Nisonger Child Behavior Rating Form	Aman et al. 1996	USA	Adaptive behaviour and behaviour that challenges	*n* = 326 Estimated IQ 55-70 3-16 years old	Teachers and parents	Rating scale	-	Available online	Yes	Teachers evaluated	Intellectual disabilities Assessing behaviour and emotional problems
PDD - Behavior Inventory	Cohen 2003	USA	Adaptive behaviour and behaviour that challenges	*n* = 311 (parent rated) *n* = 298 (teacher rated) Children attending specialist provision 1-17 years old	Teachers and parents	Rating scale	-	WPS publishing /PAR Inc.	Yes	Teacher and parent version evaluated	PDD Outcome measure for responsiveness to intervention
PDD - Behavior Inventory	Cohen et al. 2003	USA	Adaptive behaviour and behaviour that challenges	*n* = 84 Autism mild/borderline Griffith GC scores 3-6 years old	Teachers and parents	Rating scale	-	WPS publishing /PAR Inc.	Yes	Teacher and parent version evaluated	PDD Outcome measure for responsiveness to intervention
Vineland Adaptive Behaviour Scales *- Screener*	Charman et al. 2004	UK	Adaptive behaviour with optional maladaptive behaviour index	*n* = 125 Autism attending specialist provision 3-6 years old	Caregivers	Caregiver interview	New 3rd version	Pearson	Teacher version (Code B)	Teacher version (Code B)	SEN Support diagnosis and intervention planning and tracking progress

Vineland Adaptive Behaviour Scales II - *Survey Interview Form*	Chatham et al. 2018	USA/ Europe	Adaptive behaviour with optional maladaptive behaviour index	*n* = 9067 Autism 28% of sample with intellectual disabilities 1-55 years old	Teachers and parents	Semi structured interview/ rating scale	New 3rd version	Pearson	Teacher version (Code B)	Teacher version (Code B)	SEN Support diagnosis and intervention planning and tracking progress
Vineland Adaptive Behaviour Scales	Harris et al. 1995	USA	Adaptive behaviour with optional maladaptive behaviour	*n* = 20 Autism attending specialised classes 2-5 years old	Parents	Semi structured interview/ rating scale	New 3rd version	Pearson	Teacher version (Code B)	Teacher version (Code B)	SEN Support diagnosis and intervention planning and tracking progress
Vineland Adaptive Behaviour Scales II *- Parent and teacher forms*	Lane et al. 2013	Aus	Adaptive behaviour with optional maladaptive behaviour index	*n* = 39 Autism in early intervention 2-5 years old	Teachers and parents	Rating scale	New 3rd version	Pearson	Teacher version (Code B)	Teacher (and parent) version evaluated	SEN Support diagnosis and intervention planning and tracking progress
Vineland Adaptive Behaviour Scales *- Survey interview*	Perry and Factor 1989	Canada	Adaptive behaviour with optional maladaptive behaviour index	*n* = 15 Autism IQ 18-82 8-18 years old	Parent or support worker	Semi structured interview	New 3rd version	Pearson	Teacher version (Code B)	Teacher version (Code B)	SEN Support diagnosis and intervention planning and tracking progress
Vineland Adaptive Behaviour Scales *- Survey interview and teacher rating scales*	Voelker et al. 2000	Canada	Adaptive behaviour with optional maladaptive behaviour index	*n* = 90 Various disabilities 3-12 years old	Teachers and parents	Survey interview (parents) and rating scale (teachers)	New 3rd version	Pearson	Teacher version (Code B)	Teacher (and parent) version evaluated	SEN Support diagnosis and intervention planning and tracking progress
Vineland Adaptive Behaviour Scales II *- Classroom version*	Wells et al. 2009	Canada	Adaptive behaviour with optional maladaptive behaviour index	*n* = 50 Autism and intellectual disabilities 6-30 years old	Staff	Rating scale	New 3rd version	Pearson	Teacher version (Code B)	Classroom version used	SEN Support diagnosis and intervention planning and tracking progress
Wider Outcomes Survey for Teachers	Wigelsworth et al. 2015	UK	Wider outcomes with behaviour difficulties,	*n* = 6164 SEND statements or in	Teachers	Questionnaire	-	Article	Yes	Teachers evaluated	SEND Outcome measure

			relationships, exposure to bullying	receipt of SEND provision 5-15 years old							
Adaptive Behaviour Assessments											
Assessment	**Study/ Year**	**Country**	**Category**	**Sample**	**Respondents**	**Assessment Type**	**Newest/ Updated version?**	**Available?**	**Appropriate for Teachers?**	**Appropriate for School?**	**Intended population and purpose**
Adaptive Behavior Assessment System – II	Aricak and Oakland 2010	USA	Adaptive behaviour	*n* = 1690 Typically developing, intellectual disabilities, autism 5-21 years old	Teachers or family members	Rating scale	ABAS III	Pearson	CLR2	School version/teacher form used	Intellectual disabilities Screening and intervention planning
Behavior Rating Inventory for the Retarded	Sparrow and Cicchetti 1978	USA?	Adaptive behaviour	*n* = 45 Intellectual disabilities IQ 2-53 6-15 years	Staff	Interview	Unknown	Unknown	No mention of teachers	No mention of school	Intellectual disabilities Measuring behaviour
Children's Adaptive Behavior Scales	Kicklighter and Bailey 1980	USA	Adaptive behaviour	*n* = 120 IQ 70-89 and IQ 50-69 7-10 years old	No mention	Direct assessment	Unknown	Not in print? Publishers 1920-2008	Mention of teachers and assessment being "educationally useful"	Mention of teachers and assessment being "educationally useful"	Intellectual/ developmental disabilities Screening and measuring behaviour
Great Outcomes for Kids Impacted by Severe Developmental Disabilities *- Brief Adaptive Scale*	Pan et al. 2019	UK	Adaptive skills/ behaviour	*n* = 361 Special school pupils 38% autism 4-19 years old	Caregivers/ teachers	Report scale	Yes	Yes	Yes	Yes	Severe learning disabilities Research purposes, not recommended for clinical use
Great Outcomes for Kids Impacted by Severe Developmental Disabilities *- Brief Adaptive Scale*	Perry et al. 2015	Canada	Adaptive skills/ behaviour	*n* = 432 parents of children with intellectual disabilities 3-20 years	parents	Report scale	Yes	Yes	Suggests not	Suggests not	Severe learning disabilities Research purposes, not recommended for clinical use

Independent Behaviour Assessment Scale	Munir et al. 1999	Banglad esh	Adaptive behaviour	Stage 2 *n* = 1404 (typically developing) *n* = 22 (intellectual disabilities) 2-9 years old	Trained psychologists conducted assessment	Direct observation supplemented by interview	Unknown	Author	Mentions teacher manual	Mentions teacher manual - also mentions training required	Disabilities Screening
Minnesota Developmental Programming System Behavioural Scales *– Alternate Form C*	Silverman et al. 1983	USA	Adaptive behaviour	*n* = 3487 Varying levels of intellectual disabilities 0-70 years old	No info	Rating scale	Unknown	Article	No mention	No mention	Severe intellectual disabilities Diagnostic and intervention planning
Pediatric Evaluation of Disability Inventory *– Computer Adaptive Test*	Dumas et al. 2010	USA	Adaptive behaviour	Focus group *n* = 6 parents of children with disabilities *n* = 4-5 hospital clinicians Cognitive Interviews *n* = 11 parents of children with disabilities	Parents/ clinicians	Computer adaptive test	-	Website	Yes	Yes	Identifying delay, assessing improvement, monitoring and evaluating interventions Autism module in development
Pediatric Evaluation of Disability Inventory *– Computer Adaptive Test*	Kramer et al. 2012	USA	Adaptive behaviour	*n* = 20 autism professionals (incl. 2 teachers) and 18 parents of 21 autistic children 3-17 years old	Professionals /parents	Computer adaptive test	-	Website	Yes	Yes	Identifying delay, assessing improvement, monitoring and evaluating interventions Autism module in development
Pediatric Evaluation of Disability Inventory *-Patient Reported Outcome*	Kramer and Schwartz 2018	USA	Adaptive behaviour	Phase I: *n* = 8 (including 6 with intellectual disabilities) 14-21 years old Phase II: *n* = 43	Self-report scale	Self-report	Continuing development	Continuing development	n/a	Mention of schools	Disabilities Planning and evaluating outcomes of education and healthcare

				Developmental/ intellectual disabilities including autism 15-21 years old							
Pediatric Evaluation of Disability Inventory *- Patient Reported Outcome*	Kramer and Schwartz 2017	USA	Adaptive behaviour	*n* = 37 68% with intellectual disabilities, some with autism 8-21 years old	Self-report	Self-report	Continuing development	Continuing development	n/a	Mention of schools	Disabilities Planning and evaluating outcomes of education and healthcare
School Function Assessment	Coster et al. 1999	USA	Adaptive behaviour	*n* = 266 Typically developing and special needs 5-14 years old	Teacher or therapist	Questionnaire	-	Pearson	Yes	Yes	Disabilities Screening, planning, identifying and reporting progress
School Function Assessment	Davies et al. 2004	USA	Adaptive behaviour	*n* = 11 (autism) *n* = 15 (learning disabilities) 6-11 years old	Teachers	Questionnaire	-	Pearson	Yes	Yes	Disabilities Screening, planning, identifying and reporting progress
School Function Assessment	Hwang et al. 2002	USA	Adaptive behaviour	*n* = 29 (typically developing) *n* = 18 (learning disabilities) *n* = 17 (cerebral palsy) 6-14 years old	Teachers	Questionnaire	-	Pearson	Yes	Yes	Disabilities Screening, planning, identifying and reporting progress
Street Survival Skills Questionnaire	Janniro et al. 1994	USA	Adaptive behaviour	*n* = 18 Mean IQ = 45 9-17 years old	No info	Direct assessment	-	McCarron Dial Systems	No mention	No mention but participants noted as students attending public schools	Intellectual disabilities Screening, diagnosis and intervention planning

Assessments of Behaviour that Challenges											
Assessment	**Study/ Year**	**Country**	**Category**	**Sample**	**Respondents**	**Assessment Type**	**Newest/ Updated version?**	**Available?**	**Appropriate for Teachers?**	**Appropriate for School?**	**Intended population and purpose**
Aberrant Behavior Checklist	Brown et al. 2002	USA	Behaviour that challenges	*n* = 601 Intellectual disabilities 6-22 years old	Parents	Checklist	2nd Edition	Stoeltling	Has been evaluated with teachers	Has been evaluated with teachers	Intellectual disabilities Outcome measure
Aberrant Behavior Checklist	Kaat et al. 2014	USA and Canada	Behaviour that challenges	*n* = 1130 (calibration sample) *n* = 763 (validation sample) Autism, 47% with intellectual disabilities 2-18 years old	Parents	Checklist	2nd Edition	Stoeltling	Has been evaluated with teachers	Has been evaluated with teachers	Intellectual disabilities Outcome measure
Aberrant Behavior Checklist	Marshburn and Amman 1992	USA	Behaviour that challenges	*n* = 666 Intellectual disabilities 6-21 years old	Teachers	Checklist	2nd Edition	Stoeltling	Teachers used as respondents	Teachers used as respondents	Intellectual disabilities Outcome measure
Aberrant Behavior Checklist *- Irritability Subscale*	Siegel et al. 2013	USA	Irritability	*n* = 14 Autism and/or intellectual disabilities 5-18 years old	Parents	Telephone administration and checklist	2nd Edition	Stoeltling	Has been evaluated with teachers	Has been evaluated with teachers	Intellectual disabilities Outcome measure
Challenging Behaviour Interview	Oliver et al. 2003	UK	Severity of behaviour that challenges	*n* = 47 Severe intellectual disabilities 4-12 years old	Teachers for child sample	Interview	-	Author	Teacher respondents - suggests training to administer	Teachers respondents - suggests training to administer	Intellectual disabilities Outcome measure of behaviour that challenges
Developmentally Delayed Children's Behaviour Checklist	Einfeld and Tonge 1991	Aus	Behavioural and emotional problems	*n* = 5000 Intellectual disabilities unknown age	Parents, care workers, teachers, teacher aides	Questionnaire/ rating scale	Unknown	Unknown	Teacher version but Level C	Teacher version evaluated	Intellectual /developmental disabilities Supports clinical assessment and progress

Home Situations Questionnaire *- PDD*	Chowdhury et al. 2010	USA	Severity of disruptive behaviour and behavioural non compliance	*n* = 124 PDD/autism IQ > 35, some in special education 4-13 years old	parents	Rating scale	Also HSQ-ASD	Article	Items relevant to home but some may be relevant to school	School Situations Questionnaire available	PDD Outcome measure
Assessments of Autism-Related Behaviour											
Assessment	**Study/ Year**	**Country**	**Category**	**Sample**	**Respondents**	**Assessment Type**	**Newest/ Updated version?**	**Available?**	**Appropriate for Teachers?**	**Appropriate for School?**	**Intended population and purpose**
Autism Behavior Inventory	Bangerter et al. 2017	USA	Core/associated symptoms in autism	Phase 1: *n* = 353 Autism – varying communication abilities 3-21 years old Phase 2: *n* = 23 Autism 3-18+ years old	Caregivers	Rating scales (web based)	-	Article and JanssenMD	No mention of teachers/use in schools	No mention of teachers/use in schools	Autism Designed to assess and capture change
Autism Impact Measure	Kanne et al. 2014	USA	Core autism symptoms	*n* = 440 Autism IQ 45-148 2-17 years old	Caregivers	Rating scale	-	Soon to be available from WPS	No mention of teachers/use in schools	No mention of teachers/schools but author correspondence suggested may be useful Teacher version in development	Autism Outcome measure
Autism Treatment Evaluation Checklist	Charman et al. 2004	UK	Autism symptoms	Cohort 1 – *n* = 79 Cohort 2 – *n* = 55 *n* = 57 - 1 year follow up Autism provision < 6 years	Parents	Short checklist	-	Autism Research Institute website	Yes	Completed by parents, teachers or caregivers	Autism Outcome measure
Autism Treatment Evaluation Checklist	Magiati et al. 2011	UK	Autism symptoms	*n* = 22 parents of children with autism, various IQs, early intervention FU1 - 2 years later (age 4-6 years old)	Parents and teachers	Short checklist	-	Autism Research Institute website	Yes	Teachers evaluated	Autism Outcome measure

				FU2 –5-6 years later (age 9-12 years old)							
Repetitive Behavior Scale - Revised	Lam and Aman 2007	USA	Severity of restricted and repetitive behaviours	*n* = 307 Autism, some at SEN school 3-48 years old	Caregivers and staff	Rating scale	-	Unknown	Teacher respondents used in studies elsewhere	Teacher respondents used in studies elsewhere	Autism Assessing a variety of RRBs
Teacher Autism Progress Scale	Dang et al. 2017	USA	Outcomes for children on autism spectrum	*n* = 30 Autism 6-24 years old	Teachers	Rating scale	-	Article	Yes	Designed for and evaluated with teachers	Autism Outcome measure
Sensory Behaviour Questionnaire	Neil et al. 2017	UK	Sensory behaviours for moderate-severe intellectual disabilities	*n* = 66 (autism) *n* = 70 (typically developing) IQ 57-130 6-17 years old	Parents	Questionnaire	-	Freely available	May be 'useful' for educators	May be 'useful' for educators	Autism Frequency and impact of sensory behaviours

Measurement properties of the majority of the assessments (n = 16) were reported in a single study. Seven studies evaluated versions of the Vineland Adaptive Behavior Scales (VABS) (e.g. Charman et al., 2004; Harris et al., 1995). Four studies considered the Aberrant Behavior Checklist (ABC) (e.g. Brown et al., 2002; Kaat et al., 2014; Siegel et al., 2013) and current or previous versions of AAMR Adaptive Behavior Scale – II (AAMR ABS-II) (e.g. Mayfield et al., 1984; Spreat, 1982; Wells et al., 2009). Three studies evaluated the SFA (Coster et al., 1999; Davies et al., 2004; Hwang et al., 2002) and six assessments had two studies evaluating them (e.g. Pervasive Developmental Disorder – Behavior Inventory; Autism Treatment Evaluation Checklist). Four studies (Charman et al., 2004; Lane et al., 2013; Perry and Factor, 1989; Wells et al., 2009) reported on more than one assessment.

Twenty-seven of the 42 included studies were carried out in the United States (64%). Six studies were completed in the United Kingdom, four studies in Canada, two studies in Australia and one study was conducted in Bangladesh. The final two studies were conducted across two or more countries with Chatham et al. (2018) conducted in both the United States and Europe and Kaat et al. (2014) conducted in the United States and Canada. The year of publication of the included studies ranged from 1978 to 2018 with 19 of the studies conducted since 2010.

The focus of the included assessments fell into four categories: adaptive behaviour including adaptive functioning (n = 10), behaviour that challenges (n = 4), autism-related behaviour (n = 6) and both adaptive behaviour and behaviour that challenges (n = 6).

Even though studies were only included if the assessment was evaluated as an outcome measure, the Great Outcomes for Kids Impacted by Severe Developmental Disabilities - Brief Adaptive Scale (GO4KIDDS) was developed specifically for research purposes (Perry et al., 2015) and the Independent Behaviour Assessment Scale (IBAS) was developed for diagnostic or screening purposes (Munir et al., 1999). A further 10 assessments were reported to be useful for both diagnosis or screening and as outcome measures (e.g. Adaptive Behavior Assessment System - II; VABS-II) while the remaining 14 assessments were developed specifically as outcome measures (e.g. Autism Treatment Evaluation Checklist; Challenging Behaviour Interview; Teacher Autism Progress Scale). It is interesting to note that over 40% of included assessments were considered useful for screening or diagnosis.

4.3.2 Domains

The 10 adaptive behaviour assessments considered a number of different areas of functioning including social skills, communication, independence, self-help and physical skills. The Pediatric Evaluation of Disability Inventory - Computer Adaptive Test (PEDI-CAT), Pediatric Evaluation of Disability Inventory – Patient Reported Outcome (PEDI-PRO) and SFA, although addressing elements of participation, focused on functional skills of children in schools and therefore were included in the adaptive behaviour category. The Minnesota Developmental Programming System Behavioural Scales – Alternate Form C (MDPS–C) also included a domain labelled "eating behaviors" (Silverman et al., 1983, p.170). The Street Survival Skills Questionnaire (SSSQ) included skills relevant to teenagers and adolescents, for example health and safety, public services, and time, money and measurement (Janniro et al., 1994).

Four assessments focused on behaviour that challenges; the ABC (e.g. Brown et al., 2002; Marshburn and Aman, 1992), Challenging Behaviour Interview (CBI) (Oliver et al., 2003), Developmentally Delayed Children's Behaviour Checklist (DDCBCL) (Einfeld and Tonge, 1991) and the HSQ-PDD (Chowdhury et al., 2010). These assessments measured behaviour such as physical aggression and non-compliance. The HSQ–PDD and CBI both addressed aspects of severity of behaviour that challenges (Chowdhury et al., 2010; Oliver et al., 2003). The DDCBCL yielded scores relating to 'deviant' behaviour and distress to carers (Einfeld and Tonge, 1991).

Six assessments considered both adaptive behaviour and behaviour that challenges including the AAMR ABS-II (Wells et al., 2009 plus previous versions from Perry and Factor, 1989; Mayfield et al., 1984; Spreat, 1982), BASC-2 (Ellison et al., 2016; Lane et al., 2013), Pervasive Developmental Disorder – Behavior Inventory (PDD-BI) (Cohen et al., 2003; Cohen, 2003), VABS (Charman et al., 2004; Chatham et al., 2018; Harris et al., 1995; Lane et al., 2013; Perry and Factor, 1989; Wells et al., 2009; Voelker et al., 2000), Nisonger Child Behavior Rating Form (NCBRF) (Aman et al., 1996) and the Wider Outcomes Survey for Teachers (WOST) (Wigelsworth et al., 2015). Assessments such as the BASC-2 and VABS considered a wide variety of adaptive behaviour and behaviour that challenges. The PDD-BI was developed specifically for use with children on the autism spectrum and included both autism-specific and

broader, more generic skills and behaviours (Cohen et al., 2003). The WOST assessed behaviour difficulties, social relationships and experiences of bullying.

Six assessments measured autism-related behaviour and these included the Autism Behavior Inventory (ABI) (Bangerter et al., 2017), Autism Impact Measure (AIM) (Kanne et al., 2014), ATEC (Charman et al., 2004; Magiati et al., 2011), Repetitive Behavior Scale – Revised (RBS-R) (Lam and Aman, 2007), Teacher Autism Progress Scale (TAPS) (Dang et al., 2017) and the Sensory Behaviour Questionnaire (SBQ) (Neil et al., 2017). The ABI, AIM, ATEC and the TAPS all took account of the assessments' abilities to capture progress and change either during development or in the evaluation (Bangerter et al., 2017; Kanne et al., 2014; Charman et al., 2004; Dang et al., 2017). The RBS-R was devised specifically to assess RRBs in individuals with autism and also suggested potential usefulness in measuring intervention outcomes (Lam and Aman, 2007). The SBQ assessed the frequency and impact of 25 different sensory behaviours and Neil et al. (2017) suggested it may be useful in measuring outcomes.

4.3.3 Samples

4.3.3.1 Diagnosis

As per the inclusion criteria, all studies included at least some participants with intellectual disabilities. Eight assessments were devised specifically for autism including pervasive developmental disorders (e.g. PDD-BI; AIM; ATEC). A number of other assessments were described as appropriate for a variety of intellectual disabilities or developmental disabilities including autism (e.g. BASC-II; NCBRF; VABS-II).

The numbers or percentage of autistic participants or participants with intellectual disabilities varied; in two studies all or nearly all participants had intellectual disabilities and coexisting autism (e.g. Wells et al., 2009), while in other studies only some of the sample had intellectual disabilities (e.g. Ellison et al., 2016). Some studies reported the Full-Scale IQ of participants but nine studies were included on the basis that it was described or inferred that some or all of the participants needed educational provision above that which could be provided by a mainstream school (e.g. Bangerter et al., 2017; Hwang et al., 2002; Wigelsworth et al., 2015). Some studies specified the numbers of participants with each diagnosis (e.g. Chatham et al., 2018) while others did not provide the exact number of participants with intellectual

disabilities within their sample (e.g. Kanne et al., 2014). The samples will be taken into account when discussing the appropriateness of the assessments in the discussion section of this chapter.

4.3.3.2 Age

Five studies used samples six years old or younger (e.g. Charman et al., 2004; Cohen, 2003) while three studies used primary school aged samples (e.g. Aricak and Oakland, 2010; Davies et al., 2004) and Munir et al. (1999) included participants aged 2-9 years. Most other studies used samples of children and adolescents spanning school age (n = 12), children up to age 18 (n = 7) or a broad age range that included children and adults (n = 14). Where samples were split into children and adults (e.g. Oliver et al. 2003) only results from the child sample were considered in this review.

4.3.4 Methods of Assessments

As would be expected, assessment methods of the measures varied. This included direct assessment or observation (e.g. Children's Adaptive Behavior Scale; SSSQ), interviews with parents, caregivers or teachers (e.g. CBI; VABS-II) or a mixture of methods (e.g. observation and interview in the IBAS). The majority of measures were rating scales, checklists or questionnaires filled out by professionals or parents (e.g. ABC; PDD-BI; RBS-R) with only one self-report measure (PEDI-PRO).

4.3.5 Use by Teachers in an Educational Setting

Sixteen assessments were either designed specifically to be used by teachers, developed for use in schools or were evaluated using teacher respondents in the included studies. One of these studies, Oliver et al. (2003), used teachers as respondents for the CBI but did not mention whether teachers are also able to conduct the interview or if a trained interviewer is required. The HSQ–PDD has a school form of its original version available (the School Situations Questionnaire) but studies reporting on this version were not found in the systematic search. Studies which evaluated the Children's Adaptive Behavior Scales (CABS) and the IBAS mentioned or implied possible use by teachers, although teachers were not used in the studies as respondents. Kicklighter and Bailey (1980), for example, suggested that the CABS may be "educationally useful" (p.169) and Munir et al. (1999) mentioned a "teacher's manual", although

there was also suggestion that those administering the assessment in the study received "extensive training" (p.246). Studies which discussed seven of the assessments did not mention or imply the possibility of use in schools or by teachers (e.g. BRIR; MDPS – Alternate Form C; ABI). Whilst this does not mean that these assessments may not be useful with teachers, it is likely that a number of these were designed for clinical use and have not been considered or evaluated with teachers as respondents in an educational setting. Although no mention of use by teachers, the main author of the Autism Impact Measure has informed that a teacher version of the Autism Impact Measure is in development (S. Kanne, personal communication, December 2018).

Fourteen of the assessments were evaluated using teacher respondents, however four of these are subject to qualification codes upon purchase which may restrict access or use by an ordinary special needs teacher in the United Kingdom. The eligibility criteria of this review excluded studies which evaluated assessments with qualification codes restricting assessments which can only be used by qualified psychologists, however some publishers have intermediary qualification codes which require a teaching qualification along with a master's degree or further qualification in assessment in order to purchase and use specific assessments (VABS-II, ABAS-II, BASC-2 and SFA). A master's degree or further training in special educational needs is a necessary to teach this population in some U.S. states but this is not a requirement for special needs teachers in the United Kingdom. This qualification code requirement may therefore restrict the ability of the majority of special needs teachers to freely access and use these assessments without further supervision.

4.3.6 Availability and Year of Study

As a date limit was not specified for inclusion, 13 studies (30%) were conducted prior to 2000. Some of these studies may have used methods which have since been revised and updated and this must be a consideration when judging the evidence and potential uses of these assessments. In addition, older assessments may not comprehensively address adaptive behaviour involving modern technologies (Floyd et al., 2015). Furthermore, current information on some assessments proved difficult to find. Two assessments were out of print or appeared unavailable from publishers (AAMR ABS-II; CABS). Four assessments have more recent or updated versions than those considered in the included studies (ABAS-III; ABC-2;

BASC-3; VABS-III). Eight assessments (or their most recent versions) are available from publishers (ABAS-II; ABC-2; ABI; BASC-3; PDD-BI; SFA; SSSQ; VABS-III). The ATEC is available from the Autism Research Institute website and the PEDI-CAT and PEDI-PRO are available directly from their respective websites or universities, although the PEDI-PRO is still under continuing development. Nine assessments were included in the article or suggest they are available from the author (CBI; GO4KIDDS; HSQ–PDD; IBAS; MDPS–C; NCBRF; SBQ; TAPS; WOST). No information could be found in regards to accessing three of the assessments and/or there were no replies when the authors were contacted (Behavior Rating Inventory for the Retarded; DDCBCL; RBS-R). The AIM is soon to be available for purchase.

4.3.7 Measurement Properties and Quality Assessment

4.3.7.1 Content Validity

The COSMIN manual considers content validity to be "the most important measurement property" (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018; p.36) and it is relevant for the purposes of this review that assessments used by teachers were developed with teachers' input. Many of the included studies briefly described content validity (e.g. Aman et al., 1996; Kanne et al., 2014) but only four studies discussed assessment development or content validity in sufficient detail to be rated here (Dumas et al., 2010; Kramer et al., 2012; Kramer and Schwartz, 2017; Kramer and Schwartz, 2018). Of the studies providing only brief descriptions of content validity, it is interesting to note that many studies described input at the initial stages of development from the research team or clinicians such as psychologists and speech therapists but only Wigelsworth et al. (2015) and Dang et al. (2017) reported input from teachers during development.

Kramer et al. (2012) reported overall sufficient evidence for content validity of the PEDI-CAT with very good and adequate quality evidence. Dumas et al. (2010) also found sufficient content validity during assessment development. The methodological quality of the comprehensiveness study was rated as inadequate, however, due to unclear data analysis procedures and for not including a range of professionals in the sample. Kramer and Schwartz (2017) and (2018) both showed sufficient content validity of the PEDI-PRO but the evidence was rated doubtful due to lack of information on the skill or experience of the moderators.

4.3.7.2 Assessments with Sufficient Psychometric Properties

A number of assessments provided sufficient evidence for multiple measurement properties. The HSQ-PDD had good responsiveness and good internal consistency with very good quality ratings (Chowdhury et al., 2010). Construct validity was sufficient but with evidence rated inadequate due to unsatisfactory sample size. Correlations with subscales of the ABC were modest to moderate and significant. Correlations with VABS subscales were inverse but non-significant (Chowdhury et al. 2010).

Wigelsworth et al. (2015) showed very good quality evidence of sufficient internal consistency of the WOST. Structural validity was deemed insufficient according to the COSMIN criteria even though the model was reported as close to ideal fit (comparative fit index = .858, Tucker-Lewis coefficient = .838).

GO4KIDDS showed very good quality evidence of internal consistency and convergent validity with the VABS and the Scales of Independent Behaviour – Revised (Pan et al., 2019; Perry et al., 2015). Pan et al. (2019) found one principal component measuring adaptive behaviour with adequate quality evidence.

Magiati et al. (2011) provided very good quality evidence of sufficient internal consistency in all subscales of the ATEC, initially and on both follow up periods. ATEC total scores significantly correlated with the ADI-R total raw score and inversely with the VABS Composite age equivalent score at both follow up periods. These correlations, however, became insufficient by COSMIN standards when controlling for IQ. Total ATEC scores remained stable over time with large individual differences, however the methodological quality for responsiveness was rated as inadequate. Charman et al. (2004) also evaluated responsiveness and reported change in one of the three ATEC subscales but this evidence was also rated inadequate.

The TAPS measured improvements over time as compared to the ABC and Social Responsiveness Scale with adequate evidence (Dang et al., 2017). Only three other studies evaluated responsiveness of assessments. Charman et al. (2004) did not find significant change in Adaptive Behaviour Composite Score of the VABS – screener over time, while Harris et al. (1995) showed significant change of the VABS – Survey Interview Form at the first follow up but not the second. The methodological quality of these two evaluations was rated inadequate

due to the COSMIN manual considering paired t-tests an inappropriate measure of responsiveness.

The PDD-BI showed varied data on IRR and convergent validity; teacher IRR across subscales ranged from moderate to high (range .55 to .93) and was more strongly correlated than parent-teacher IRR. Test-retest reliability for the teacher scale was sufficient with very good quality evidence (range .73 to .97). Convergent validity with the Childhood Autism Rating Scale was moderate but significant (.50), NCBRF was low to moderate (range .16 to .66) and VABS subscales significant with a range from .31 to .81.

Lam and Aman (2007) provided very good quality evidence of sufficient internal consistency of the RBS-R. IRR for the different subscales ranged from .57 to .73 for the younger sample and -.24 to .95 for the older sample. A five-factor solution for the RBS-R was adopted from the EFA which accounted for 47.5% of the variance, below the cutoff of 50% for good structural validity. Adequate evidence was found of a close to ideal fit with a root mean square error of approximation (RMSEA) of .061, just outside of the COSMIN level for sufficient structural validity.

The SFA showed moderate to good convergent validity with the VABS - Classroom version for the learning disabilities group with very good quality evidence (Hwang et al., 2002). Davies et al. (2004) found sufficient IRR between teachers and therapists for only two of the three scales. A two-factor solution was indicated by Coster et al. (1999) although this evidence was of inadequate quality due to a small sample size.

Kaat et al. (2014) provided very good quality evidence for sufficient internal consistency of the ABC and appropriate convergent and divergent validity with the CBCL and VABS. Construct validity varied across studies. An EFA by Marshburn and Aman (1992) found that a four-factor solution accounted for 52% of the variance. A confirmatory factor analysis (CFA) of a five-factor solution by Brown et al. (2002) yielded a sufficient RMSEA according to COSMIN criteria (< 0.06), however this threshold was not reached by Kaat et al. (2014). Siegel et al. (2013) found very good quality evidence for no significant difference between written and telephone administrated ABC scores.

The VABS showed reasonable convergent validity with the AAMD Adaptive Behavior Scale (AAMD ABS) (Perry and Factory, 1989) and a significant moderate inverse correlation with the CARS (Wells et al., 2009).

4.3.7.3 Teacher and Parent Rating Scales

An additional evaluation of assessments included in this review considered correlations or comparisons between teacher and parent responses. Lane et al. (2013) found significant differences on the parent-teacher correlations of the adaptive skills composite of the BASC-2 with teacher ratings significantly less severe than parent ratings. Similarly, Ellison et al. (2016) found that parent ratings were significantly more severe than teacher ratings on the Adaptive Skills Composite and all three adaptive subscales of the BASC-2. Lane et al. (2013) also showed that parent-teacher correlations on VABS-II domains were all significant with no significant differences. These studies were of very good methodological quality. Voelker et al. (2000) compared parent-teacher ratings on the VABS and found that correlations between teacher and parent scores were high for the summary score and all domains apart from the socialisation domain. Very good quality evidence showed, again, that parents consistently and significantly reported lower adaptive behaviour skills than teachers. However, when the 169 overlapping items from VABS classroom and survey form were analysed for IRR, parents reported higher skill level on 70% of comparisons with 93% of correlations significant. This evidence was considered indeterminate and of inadequate quality due to the use of the phi correlation coefficient. Aman et al. (1996) considered teacher and parent agreement on the NCBRF. They found that correlations were significant but ranged from .22 to .54, indicating differences between teacher and parent ratings of a child's adaptive skills or 'problem' behaviour.

4.3.7.4 Other Assessments

Twenty-one of the included studies only reported on one measurement property for the sample of interest. For seven of the 26 assessments there was information on only one measurement property from only one study (e.g. ABAS-II; CABS; IBAS; MDPS – C). Most other assessments had three or more measurement properties evaluated. Neil et al. (2017) found good internal consistency for the SBQ and convergent validity with the Short Sensory Profile. Spreat (1982) found no significant differences between weighted and non-weighted items on previous versions of the AAMD ABS and Chatham et al. (2018) estimated minimal clinically

important differences of the Composite Score of the VABS to be 2-2.5 points for the relevant sample.

Of the studies conducted before the year 2000, it was noted that six studies had inadequate ratings for one or more measurement property (Aman et al., 1996; Coster et al., 1999; Kicklighter and Bailey, 1980; Harris et al., 1995; Mayfield et al., 1984; Sparrow and Cicchetti, 1978). Weaknesses shown in older scales may be an indication of progress made in scale development and validation over time and/or improved reporting within peer reviewed studies (Floyd et al., 2015). The evaluation of measurement properties and the quality assessment of each study are summarised in Table 4.

Table 4. *Risk of Bias and Measurement Properties*

Assessments of Adaptive Behaviour and Behaviour that Challenges											
Assessment	**Author**	**Content Validity**				**Construct Validity**		**Reliability**			**Responsiveness**
		Development	**Parents/Experts**			**Structural Validity**	**Hypothesis Testing (convergent)**	**Internal Consistency**	**Inter-rater reliability**	**Test-retest reliability**	
		Prom Design	**Relevance**	**Comprehensiveness**	**Comprehensibility**						
AAMD Adaptive Behaviour Scale - Public School Version	Mayfield et al. 1984								Inadequate -	Inadequate +	
AAMD Adaptive Behaviour Scale - School Edition	Perry and Factor 1989						Very good +			Doubtful +	
AAMD Adaptive Behaviour Scale - Part 2	Spreat 1982						Adequate +				
AAMR Adaptive Behavior Scales - School 2nd Edition (revision of AAMD school edition)	Wells et al. 2009						Adequate +				
Behavior Assessment System for Children - 2	Ellison et al. 2016						Very good +				
Behavior Assessment System for Children - 2	Lane et al. 2013						Very Good +				
Nisonger Child Behavior Rating Form	Aman et al. 1996					Inadequate -	Adequate +		Inadequate -		

PDD - Behavior Inventory	Cohen 2003	Inadequate **?**		Very Good +	Adequate + -		
PDD - Behavior Inventory	Cohen et al. 2003		Adequate +				
Vineland Adaptive Behaviour Scales II *- Screener*	Charman et al. 2004						Inadequate + -
Vineland Adaptive Behaviour Scales II *- Survey Interview Form*	Chatham et al. 2018		Very good +				
Vineland Adaptive Behaviour Scales	Harris et al. 1995						Inadequate + -
Vineland Adaptive Behaviour Scales II *- Parent and teacher forms*	Lane et al. 2013		Very good +				
Vineland Adaptive Behaviour Scales *- Survey interview*	Perry and Factor 1989		Adequate +				
Vineland Adaptive Behaviour Scales II *- Classroom version*	Wells et al. 2009		Adequate +				
Vineland Adaptive Behaviour Scales	Voelker et al. 2000		Very good +		Inadequate **?**		
Wider Outcomes Survey for Teachers	Wigelsworth, et al. 2015	Very good -		Very good +			

Adaptive Behaviour Assessments											
Assessment	**Author**	**Content Validity**				**Construct Validity**		**Reliability**			**Responsiveness**
		Development	**Parents/Experts**			**Structural Validity**	**Hypothesis Testing (convergent)**	**Internal Consistency**	**Inter-rater reliability**	**Test-retest reliability**	
		Prom Design	**Relevance**	**Comprehensiveness**	**Comprehensibility**						
Adaptive Behavior Assessment System	Aricak and Oakland 2010					Very good -					
Behavior Rating Inventory for the Retarded	Sparrow and Cicchetti 1978					Inadequate +	Inadequate +		Adequate +		
Children's Adaptive Behavior Scales	Kicklighter and Bailey 1980							Inadequate total score + -			
Great Outcomes for Kids Impacted by Severe Developmental Disabilities *- Brief Adaptive Scale*	Pan et al. 2019					Adequate +	Very good +	Very good +			
Great Outcomes for Kids Impacted by Severe Developmental Disabilities *- Brief Adaptive Scale*	Perry et al. 2015						Very Good +	Very good +			
Independent Behaviour Assessment Scale	Munir et al. 1999							Very good +			
Minnesota Developmental Programming System Behavioural Scales *– Alternate Form C*	Silverman et al. 1983					Adequate **?**					

Pediatric Evaluation of Disability Inventory - Computer Adaptive Test	Dumas et al. 2010	Relevance Doubtful +							
		Comprehensiveness Inadequate -							
		Comprehensibility Doubtful +							
Pediatric Evaluation of Disability Inventory - Computer Adaptive Test	Kramer et al. 2012		Adequate +	Adequate +	Adequate +				
Parents *Professionals*			Very good +	Very good +					
Pediatric Evaluation of Disability Inventory - Patient Reported Outcome	Kramer and Schwartz 2018	Relevance Very good +							
		Comprehensiveness Doubtful +							
		Comprehensibility Doubtful +							
Pediatric Evaluation of Disability Inventory - Patient Reported Outcome	Kramer and Schwartz 2017		Doubtful +		Doubtful +				
School Function Assessment	Coster et al. 1999					Inadequate +			
School Function Assessment	Davies et al. 2004								Adequate + -

School Function Assessment	Hwang et al. 2002						Very good +				
Street Survival Skills Questionnaire	Janniro et al. 1994						Doubtful + -				

Assessments of Behaviour that Challenges

Assessment	Author	Content Validity				Construct Validity		Reliability			Responsiveness
		Development	Parents/Experts			Structural Validity	Hypothesis Testing (convergent)	Internal Consistency	Inter-rater reliability	Test-retest reliability	
		Prom Design	Relevance	Comprehensiveness	Comprehensibility						
Aberrant Behavior Checklist	Brown et al. 2002					Very good -					
Aberrant Behavior Checklist	Kaat et al. 2014					Very good -	Very good +	Very good +			
Aberrant Behavior Checklist	Marshburn and Aman 1992					Adequate +					
Aberrant Behavior Checklist - *Irritability Subscale*	Siegel et al. 2013						Very good +				
Challenging Behaviour Interview	Oliver et al. 2003						Adequate +				
Developmentally Delayed Children's Behaviour Checklist	Einfeld and Tonge 1991						Adequate +	Doubtful +	doubtful teachers + teacher-TAs -	Doubtful +	
Home Situations Questionnaire - *PDD*	Chowdhury et al. 2010					Inadequate +	Very good +	Very good +			Very good +

Assessments of Autism-Related Behaviour												
Assessment	**Author**	**Content Validity**				**Construct Validity**		**Reliability**			**Responsiveness**
		Development	**Parents/Experts**			**Structural Validity**	**Hypothesis Testing (convergent)**	**Internal Consistency**	**Inter-rater reliability**	**Test-retest reliability**	
		Prom Design	**Relevance**	**Comprehensiveness**	**Comprehensibility**						
Autism Behavior Inventory	Bangerter et al. 2017					Inadequate **?**	Very good +	Very good + -		Inadequate **?**	
Autism Impact Measure	Kanne et al. 2014					Adequate +	Very good +		Doubtful + -	Doubtful + -	
Autism Treatment Evaluation Checklist	Charman et al. 2004										Inadequate -
Autism Treatment Evaluation Checklist	Magiati et al. 2011						Very good +	Very good +			Inadequate -
Repetitive Behavior Scale - Revised	Lam and Aman 2007					Adequate -		Very good +	Very good + -		
Teacher Autism Progress Scale	Dang et al. 2017										Adequate +
Sensory Behaviour Questionnaire	Neil et al. 2017						Adequate +	Very good +			

Measurement properties ratings: + sufficient; - insufficient; ? indeterminate; +- inconsistent.

Evidence quality ratings: very good, adequate, doubtful or inadequate.

4.4 Discussion

Twenty-six assessments were found with potential for use in school settings to measure progress of adaptive behaviour, behaviour that challenges or autism-related behaviour of children on the autism spectrum with coexisting intellectual disabilities. When considering the appropriateness of these assessments for use by teachers in special schools, there are a number of factors that need consideration: (a) the purpose of the assessment, (b) the usability of the assessment, e.g. whether consideration of use by teachers had been made during development, (c) the applicability of use alongside the school curriculum in the United Kingdom and (d) the measurement properties of the assessment.

4.4.1 Assessment Purpose and Intended Population

As mentioned in the results section above, even though use as an outcome measure was necessary for inclusion in this review, 42% (n = 11) of the included assessments also support screening or diagnosis with one further assessment developed specifically for research purposes. The importance of the assessment purpose was explored in Chapter 2. Assessments which either attempt to serve multiple purposes or are used for purposes for which they were not intended may be less effective at measuring for a specific purpose. As Pellegrino (2014) identified, "one type of assessment does not fit all purposes or contexts for use" and it must not be assumed that "an assessment is appropriate and interpretable for a particular context of use without determining if there is evidence regarding the validity of such assumptions within the context" (p.68). Assessments devised for screening and diagnosis will have a greater focus on areas which distinguish groups and identify difficulties rather than showing progress through changes in scores. It is necessary, therefore, to take into account the original purpose for which the assessment was developed when considering the appropriateness of an assessment in specific contexts. As recognised by Haynes et al. (1995), evaluations of properties of an assessment for one purpose are not necessarily generalisable to the use of the assessment for other purposes. As an example, GO4KIDDS showed some initial promise for use by teachers in U.K. special schools (Pan et al. 2019). As it was originally developed for research purposes and the authors specifically noted that it was not recommended for contexts other than research (Perry et al. 2015), further validation would be needed before it could be considered an appropriate measure of progress and outcomes for use in schools (McConachie et al., 2015). Assessments developed specifically to measure outcomes and progress (e.g. AIM; ATEC; CBI; PDD-BI; RBS-R; SBQ; TAPS) are likely to

be more effective, valid and reliable for this purpose than those which were developed for multiple purposes.

Another consideration is the population for which the assessment was intended and, further, the population with which the assessment has been evaluated. Eight assessments were developed specifically for use with individuals on the autism spectrum but only three of these were developed for use in schools or evaluated in the included studies using teacher respondents (PDD-BI; ATEC; TAPS). Of the 14 assessments evaluated by teachers, only Wells et al. (2009) who evaluated the AAMR ABS-II and Perry and Factor (1989) who reported on the previous AAMD version specified that all or nearly all participants were on the autism spectrum with coexisting intellectual disabilities. The ATEC, PDD-BI, SFA, TAPS and the WOST used samples who needed at least some special educational provision but the level of intellectual disabilities amongst participants is likely to have varied (e.g. Charman et al. 2004; Dang et al. 2017; Magiati et al. 2011). Considering the often-complex educational needs of this specific population as discussed in Chapter 1, it would be beneficial for further studies on these assessments to be carried out using a relevant sample.

4.4.2 Usability of Assessment by Teachers in the United Kingdom

Only 12 of the 26 assessment tools were developed specifically for use by teachers or in schools and the use of four of these assessments (ABAS-II; VABS-II; BASC-2; SFA) are subject to publishers' qualification codes which potentially limits their use. Although fewer than half of included assessments were developed for use by teachers in schools, it is encouraging that assessments are being developed specifically for this purpose and that use in an educational setting is a consideration during development. Some recently evaluated and available assessments such as the ABI, RBS-R and the SBQ may also have potential for assessing particular areas of difficulty in schools however it is necessary for more research to be conducted on using teacher respondents to further assess applicability and appropriateness for use in education settings. Similarly, the COSMIN manual considers content validity the most important measurement property and if high quality evidence suggests that the content validity is insufficient, then the other properties will not be considered (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). Only four studies provided enough information to evaluate content validity and, of the other included assessments which provided brief information about the development process, only two reported input from teachers at the development stage (TAPS; WOST). Content validity is a vital consideration and, in this

context, requires input from teachers during development in the areas identified in the COSMIN checklist including relevance of the items, comprehensiveness of the assessment and comprehensibility of the assessment instructions, items and response options.

When considering the use of assessments in schools, the assessment method may also need to be taken into account. As discussed in Chapter 2, report measures are most convenient and simple to use compared with direct assessment which may need preparation, time, a specific space or particular resources. Interviews require an interviewer and, as with direct assessment, may take time. For example, the interview time for the BRIR was 45-90 minutes (Sparrow and Cicchetti, 1978) in comparison with a number of the report measures which reported completion times of 10-20 minutes (e.g. Aricak and Oakland, 2010; Ellison et al., 2016; Einfeld and Tonge, 1991). Although some report measures may take longer to complete, they allow more flexibility than other methods as report measures do not rely upon extra resources, interaction with the child or the need for additional facilitators.

The ATEC, WOST, CBI and GO4KIDDS were the only four assessments evaluated by teacher respondents which were conducted in the United Kingdom. Although results of studies conducted in one country may be applicable to another, when considering the use of these assessments by teachers it may be useful to consider the appropriateness of these assessments in U.K. schools and their use alongside the relevant curriculum. In this case, it may be particularly beneficial for the TAPS, which was specifically devised to be used by teachers and showed responsiveness to change over time, to be evaluated within special schools in England. Similarly, although mainly involving clinicians in its development, the PEDI-CAT received some input from teachers during the development process and, with further research on its use in schools particularly in the United Kingdom, may be appropriate for use in educational settings. Initial evaluation of the PEDI-PRO suggests it may be useful as a pupil report measure but, again, further evaluation in schools, particularly in the United Kingdom is needed.

4.4.3 Measurement Properties

Although a number of assessments showed sufficient evidence for various different measurement properties, few were evaluated with a relevant sample or in an appropriate setting. The ABC and the RBS-R showed promise for use by teachers with autistic children with intellectual disabilities to assess behaviour that challenges and repetitive behaviour respectively, as did the PDD-BI. However, these assessments need further evaluation of their

responsiveness to change and their use in schools in England. Both studies evaluating the ATEC reported on the responsiveness of the assessment, which is an important measurement property to consider when determining how well an assessment measures progress. Charman et al. (2004) reported change over time on the social, language and communication subscale but no significant differences during the time period (mean 11 months) of scores on the sensory and cognitive awareness, sociability or health, physical and behaviour subscales. Magiati et al. (2011) produced similar results finding that, although over time the ATEC scores remained relatively stable, there were different individual patterns of change which may be considered in line with expectations for a heterogeneous condition such as autism. These results suggest that it is unlikely that the ATEC would be able to show progress and change over shorter periods of time (e.g. termly or half-termly) although the tool may be useful for teachers to show longer term progress. Both of these evaluations, however, were rated as having inadequate methodological quality. The WOST showed high internal consistency and the CFA indicated a close to ideal model fit (Wigelsworth et al., 2015) but further testing may be needed to determine the responsiveness of the scale. Some items on this scale are more relevant to individuals with mild intellectual disabilities rather than moderate to severe intellectual disabilities so it may be useful for further research to be conducted using this sample. It is also necessary for further studies to be carried out on the content validity of the use of these assessments with the specific population considered in this study, as mentioned above.

4.4.4 Assessments Appropriate for Schools

Twenty-six relevant assessments were identified in this review. When taking into account the factors considered above, there are few, if any, assessments which have been evidenced to be entirely appropriate for teachers to show progress of autistic pupils with intellectual disabilities without need for further evaluation. Many of the included assessments were originally developed for other purposes or are limited by qualification codes and, therefore, may need to be adapted in order to be appropriate for use in schools. Many do not have sufficient evidence of a number of robust measurement properties, particularly responsiveness, when using teacher respondents. Most have not been evaluated with individuals on the autism spectrum with intellectual disabilities when used by teachers in educational settings in the England. Considering the various factors discussed above, the ATEC and the TAPS may have potential for use in special schools to show progress of pupils on the autism spectrum as they are both autism-specific outcome measures which were

developed to measure change in broad range of behaviours relevant to autistic individuals. However, further evaluation is necessary, particularly use of the ATEC with autistic samples with intellectual disabilities in schools and evaluation of further measurement properties of the TAPS in special schools in England. The PDD-BI may also be useful with further evaluation in the United Kingdom including evaluations of the assessment's responsiveness to change. The teacher version of the Autism Impact Measure could also be a valuable addition to the pool of current available assessments upon completion. In light of the discussion and evaluation of the identified assessments, there is a clear need to develop robust assessments for use by special needs teachers to measure progress and outcomes of the specific educational needs and challenges of autistic pupils with intellectual disabilities.

4.5 Limitations

This systematic review has, to the best of the author's knowledge, been the first to consider the educational appropriateness of assessment tools which measure progress in adaptive behaviour, behaviour that challenges and autism-related behaviour for children on the autism spectrum with intellectual disabilities. It has systematically identified relevant assessments, summarised and reviewed evidence pertaining to their measurement properties and examined the assessments in respect of their use by teachers in special schools. It has also devised some adaptations to the COSMIN checklist for the purposes of teacher-reported outcomes in educational settings. This review also provides a resource for teachers which summarised the potential uses of included assessments with different samples as well as reporting on their measurement properties.

There are, however, a number of limitations of the current systematic review. Firstly, some notable assessments were not included in this study. This may be for a number of reasons. Relevant assessments may have been used in studies for diagnostic/screening purposes, to discriminate between groups or may have used a sample of individuals without intellectual disabilities and would therefore have been excluded. Some assessments were also excluded on qualification code while newer assessments and recent versions may not have been included if there have not yet been studies of their measurement properties published. Although systematic reviews reduce bias in the collection of data, they do not overcome the limitation of publication bias and may still be subject to some subjectivity in inclusion and categorisation of studies along with the interpretation of results (Garg et al., 2008). Discussions were held with the research supervisors in the attempt to reduce this bias.

Furthermore, systematic searches may result in the exclusion of some relevant articles, studies or assessments. Attempts were made to find and include relevant studies which weren't identified through the search using the ancestry method and a search of the grey literature.

Many of the included studies did not report on a wide range of measurement properties and those properties reported outside of peer-reviewed literature, for example in books, were not included. The COSMIN checklist guided the quality assessment but was adapted to suit the specific needs of the aims of this review, therefore it is necessary to interpret the methodological quality and summary of measurement properties with caution if considering the results in a broader context than is specified here.

In addition to limitations of the review methodology, there are also a number of assessment tools used in special schools in the United Kingdom which are notably missing from the literature. These include statutory assessments such as the P Scales, the Early Years Foundation Stage Profile and the new Pre-Key Stage Standards and Engagement Model. Commonly used assessment software such as B Squared assessment software was also not included. Not only did these assessments not appear in either this systematic search or further searches of peer-reviewed journals, grey literature searches specifically for these assessments also yielded no information on evaluation of their properties. B Squared was also contacted for information on their measurement properties but no reply was received. In a similar way, measurement approaches such as Goal Attainment Scaling are less likely to be included when considering evaluations of the measurement properties of these tools. McConachie et al. (2015) mentioned that criterion-referenced assessment and other assessment approaches are often not examined for their measurement properties in research. Persicke et al. (2014) recognised that, due to a lack of expertise around measurement properties, limited information on measurement properties of assessments are available in fields such as education. Teachers may intrinsically 'know' which assessments are helpful for them and their pupils and not rely on further academic evaluations of assessments which they find useful. With school wide assessment policies often chosen and developed by individuals predominantly working outside of the classroom, it is important that the gap between robust and sound assessments and their effective use by teachers is bridged.

With these limitations in mind, systematic reviews and further research replicating and evaluating the results here are recommended to address the lack of research in this area.

4.6 Conclusions and Recommendations

This systematic review has addressed the first of the two research questions by identifying assessments which can be used by special needs teachers to assess the progress and outcomes of pupils on the autism spectrum with coexisting intellectual disabilities in the areas that may constitute barriers to learning as identified in Chapter 1. The review summarised the assessment information and identified the assessment methods, previous uses of the assessments and the populations they have assessed. In addressing the second research question and determining which assessments are appropriate for these purposes, factors that were considered included the availability of the assessment, accessibility and ease of use by teachers, whether the assessments had been evaluated with a relevant population and with teacher respondents, and the outcome of the evaluation in relation to their measurement properties. Although some assessments such as the ATEC and TAPS were found to have potential for uses relevant to this thesis, this systematic review has identified a gap in the literature and shown that no robust, autism-specific outcomes assessments are currently available for teachers to use to show progress in barriers to learning for pupils with autism in special schools.

There are a number of further recommendations as a result of these findings. Firstly, many assessments used in schools have not been evaluated in peer-review literature and it is recommended that widely used assessments in special schools have their measurement properties evaluated. Secondly, as recommended by McConachie et al. (2015), it is critical that stakeholders are involved in the development of new assessments; specifically, that teachers are included in the development process of teacher assessments and that they support decisions on skills and behaviour which are most useful to assess. Thirdly, it is important for responsiveness to be evaluated including measuring small amounts of progress over shorter periods of time (e.g. termly or half-termly) for the purpose of showing progress in schools. Finally, to address the need for robust assessment tools specifically developed with and for teachers which can be used in U.K. special schools to show progress in barriers to learning for autistic children with intellectual disabilities, it is recommended that an assessment is developed specifically for this purpose with appropriate input at the development stage and evaluated with an appropriate sample in a relevant educational setting.

Chapter 5. Study 2a – Focus Groups

5.1 Introduction

The results of the systematic review indicated that there was a lack of robust assessments which could be used by teachers in special schools in the United Kingdom to identify and show progress in areas which could be considered barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities. The findings were similar to previous systematic reviews and evidenced the research gap which is the focus of this thesis. Stage 2 then consisted of further defining the construct and exploring the areas which make up barriers to learning for this group of pupils as well as considering the necessary features of the assessment tool and generating and evaluating a list of items to be included in the assessment. The input of different stakeholders was considered when this stage of the research was being designed including teachers, parents and the pupils themselves but, due to time and resource constraints, not all groups could be consulted. At this stage of the research it was felt that teachers were the crucial group who would be regularly using the assessment and interpreting the results for purposes of teaching and learning in schools. Teachers also have specialist knowledge in pedagogy and interventions and are often best placed to provide insight into both the teaching and learning needs of their pupils. Concentrating on teachers' priorities for assessment at the development stage would allow the research to remain focused. Although extremely important stakeholders in their children's education, it was felt that parents' input would be useful in future studies after the first version of the assessment had been created and piloted. Teacher focus groups, which were the first part of this stage of the research, were the main inductive element in relation to the 'barriers to learning' construct and the generation of the initial list of items.

5.1.1 Background

Although much of the research focusing on autism education is about inclusive education in mainstream schools, research specifically involving special needs teachers is becoming more common. In recent studies, for example, special needs teachers have been asked their opinions on different teaching interventions for pupils on the autism spectrum (Mills and Chapparo, 2018) and thoughts on behaviour difficulties (e.g. Adams et al., 2019; Welsh et al., 2019). Although research has explored parent priorities for teaching and outcomes for autistic children (McConachie et al. 2018), little research has been conducted

on outcomes that special needs teachers consider to be important. Azad and Mandell (2016) conducted interviews with parents and teachers of children on the autism spectrum in the United States and compared their concerns and priorities for intervention. They found that teachers' primary and secondary concerns were "problem behaviour" and "deficits in social interaction" and these were also the main concerns of parents (p.437). For teachers, these top two concerns were followed by restricted and repetitive behaviours (RRBs), communication and self-help. Teachers were least concerned about academic development (Azad and Mandell, 2016). Helps, Newsom-Davis and Callias (1999) conducted a British study which considered mainstream and special needs teachers' knowledge and understanding of autism. Teachers were asked about the difficulties they face when working with pupils on the autism spectrum, with the most cited difficulty being lack of knowledge, followed by repetitive or obsessional behaviours, poor communication and aggressive and self-injurious behaviours. The priorities identified in these studies reflect the areas considered in Chapter 1 and Chapter 4. Further studies considering special needs teachers' opinions on the important skills and progress of their pupils on the autism spectrum do not appear to have been conducted, particularly in the United Kingdom. Again, there appears to be little, if any, research considering special needs teachers' views of educational assessments.

At this stage of the research, two focus groups were conducted with the following research aims:

(a) to identify important areas of progress and barriers to learning for pupils on the autism spectrum,
(b) to consider broad areas and specific skills or behaviours which special needs teachers think are important to assess for autistic pupils and
(c) to consider the features of assessment tools which are useful to special needs teachers in the classroom.

5.2 Methods

5.2.1 Participants

A number of headteachers or members of the senior leadership team at local special schools in Hampshire were contacted and invited to participate in the focus groups. Twenty-one teachers (eighteen female and three male) from two special schools in Southern England participated. One school catered for pupils aged 2-19 years with severe or profound

intellectual disabilities and the other 3-11 years with moderate or severe intellectual disabilities. The two schools were federated in 2017 and came under the leadership of the same executive headteacher. Both schools were rated good in their most recent Ofsted inspections. The assistant headteacher from one school agreed to be involved after initial contact was made by the researcher and, as a result of positive feedback between schools after the first focus group, the assistant headteacher from the second school agreed to participate. Teacher participants were recruited through the assistant headteachers. All of the participants were qualified teachers currently working in a special school with recent experience of working with pupils on the autism spectrum with coexisting intellectual disabilities. All teachers available on the day of the first focus group participated and seven of the 10 available teachers attended and participated during the second focus group. To reduce time pressure and burden on the teacher participants, further specific data on age, race/ethnicity or previous teaching experience were not collected. Three participants were members of the schools' senior leadership teams and the other participants were classroom or specialist subject teachers. By using participants from the same school in each focus group, the research could be conducted on the school premises during working hours, minimising burden on teachers' time and workload. For the purposes of this research, it was thought that the discussions benefitted from participants being familiar with each other as they may be better able to relate to each other's contributions and feel comfortable challenging or disagreeing (Rabiee, 2004). It was also important that the participants may feel empowered when given an opportunity to discuss assessment, as assessment is usually a prescribed process for teachers (Kitzinger, 1995).

5.2.2 Ethical Compliance

Special needs teachers who had shown an interest in participating were provided with information sheets explaining the research background, the focus group procedures including audio recording, the voluntary nature of participation, the right to withdraw, confidentiality and information about data storage. Written consent was then obtained with the participants completing and signing two consents forms, one copy which was kept by the participants and the second copy which was retained by the researcher. Participants were verbally reminded of the voluntary nature of their participation and their unconditional right to withdraw during the focus groups. Participants were also verbally informed that no particular answers were expected of them and that they could give their opinions freely.

5.2.3 Procedure

Both focus groups took place on school premises during school time or staff meeting time and were moderated by the author. The moderator followed the same semi-structured focus group interview guide for both focus groups which had been refined following a pilot test (Appendix 4.1). With introductions, summaries and information also provided during this time, the discussion elements of the focus group lasted for approximately 47 and 55 minutes. Teachers were asked their opinions on important progress for this specific group of pupils, barriers to learning and their thoughts on assessment. All participants were given adequate opportunities to contribute to the discussion if they desired. The focus groups were audio recorded with notes made during the discussions in order to develop follow up question and support summaries. The moderator verbally summarised the discussion after each question and conducted participant checking. The data were transcribed orthographically including all audible spoken words, sounds, utterances, interruptions, laughter, pauses and mumbles of agreement.

5.2.4 Data Analysis

The data were analysed using template analysis. Template analysis goes further than more general thematic analysis by allowing for an *a priori* coding template which is refined as the data analysis develops (King, 2012). Template analysis often uses four or more levels of themes and subthemes, allowing the analysis to consider the depth and detail of the data, in contrast with only one or two layers of subthemes often found in thematic analysis (Brooks et al., 2015). A six-step approach to template analysis was followed as outlined by Brooks et al. (2015).

Firstly, *a priori* themes were decided upon based on the review of the literature and systematic review, the author's professional knowledge and experience as a teacher and the initial engagement with the data. During the data analysis procedure, these main themes and their subthemes were revised until a final version was decided upon and the six main themes defined. The initial coding and determination of themes was conducted by the author with discussions taking place with the doctoral supervisors until the concept tree and final themes were agreed. As well as the verbal participant checking, the final definition of themes and the concept tree outlining subthemes were sent to a participant from each focus group for confirmation that the themes accurately reflected and covered the content of the discussions.

5.3 Results

Six main themes were identified from the two focus group discussions with a total of four levels of themes and subthemes. The main themes were categorised as *Autism-related barriers and atypical skill development, Overcoming barriers, Priorities for autistic pupils, 'True mastery', Assessing the bigger picture* and *Practicalities of assessment*. The themes and subthemes are shown in Table 5.

Table 5. *Themes Concept Tree*

Theme 1. Autism-related barriers and atypical skill development
1.1 Intrinsic barriers to learning
1.1.1 Restricted and repetitive behaviours
1.1.1.1 Rigidity in thought
1.1.1.2 Repetitive and ritualistic behaviours
1.1.1.3 Restricted behaviours and interests
1.1.1.4 Difficulty with change and transitions
1.1.2 Physical and sensory needs
1.1.2.1 Sensory seeking
1.1.2.2 Sensory aversions and the environment
1.1.2.3 Basic physical needs
1.2 Atypical skill development resulting in barriers to learning
1.2.1 Emotional states
1.2.1.1 Heightened states including anxiety and frustration
1.2.1.2 Self-awareness and recognising emotions
1.2.1.3 Regulating emotions and behaviour
1.2.2 Learning behaviours
1.2.2.1 Attention, focus and readiness to learn
1.2.2.2 Engagement and enjoyment
1.2.2.3 Confidence and self-esteem
1.2.2.4 Risk taking and problem solving
1.2.3 Functional communication
1.2.3.1 Importance of the ability to communicate wants, needs and feelings
1.2.3.2 Difficulties due to lack of communication skills
Theme 2. Overcoming barriers
2.1 Pupil-teacher relationships
2.1.1 Knowing the pupil well
2.1.2 Pupil trusting the teacher
2.2 Collaboration
2.2.1 Discussions with other teachers/teaching assistants
2.2.2 Wider collaboration between schools
2.3 Links with parents/home
2.3.1 Building relationships with parents
2.3.2 An understanding of the pupil's home life
Theme 3. Priorities for autistic pupils
3.1 Generalisation to the real world
3.1.1 Generalising skills with different people
3.1.2 Generalising skills in different settings
3.1.3 Independence and self-care skills

3.1.4 Coping in the real world
3.1.5 Limitations of teaching meaningful context
3.2 Community access/engagement
3.2.1 Being accepted in the community
3.2.2 Engaging in the wider community
3.2.3 Socially acceptable behaviours
Theme 4. 'True mastery'
4.1 Rote learning vs real understanding
4.2 Assessment reflecting true ability
4.2.1 Missing details
4.2.2 Measuring breadth of learning
4.3 Do we know what true mastery is?
4.3.1 'Knowing' a pupil can do something
4.3.2 Constantly questioning and checking mastery
4.4 Interpretation can vary
4.4.1 Precision and room for variation in judgement 4.4.2 Teachers' differing opinions/values
Theme 5. Assessing the bigger picture
5.1 Looking at pupils holistically
5.2 Personalisation/individualisation
5.2.1 Heterogeneity
5.2.2 Adapting assessment for that pupil
5.3 Importance of recording the nuances and subtleties
5.3.1 Small things can be huge achievements
5.3.2 Prioritising academic progress over learning behaviours and emotional states
5.4 Regression
5.4.1 Realities of regression
5.4.1.1 It's negative
5.4.1.2 It's necessary
5.4.2 Difficulties assessing regression
Theme 6. Practicalities of assessment
6.1 Nowhere to record that
6.1.1 Some progress 'can't' be measured
6.1.2 Assessments don't allow for recognition of certain progress
6.2 Autism 'not fitting' frameworks or systems
6.2.1 Non-linear progress and spiky profiles
6.2.2 Maintaining skills and gaps in learning
6.3 Helpful aspects of assessment
6.3.1 Next steps
6.3.2 Smaller steps

5.3.1 Theme 1: "It's a bit of a hindrance to his work" - Autism-Related Barriers and Atypical Skill Development

This theme incorporated the areas that teachers identified as barriers to learning for pupils on the autism spectrum along with areas which teachers felt were crucial to address

before curriculum learning or more complex skills could be considered. It was divided into two subthemes; *intrinsic barriers to learning* and *atypical skill development resulting in barriers to learning*.

5.3.1.1 Intrinsic Barriers to Learning

Teachers identified intrinsic barriers to learning as those related to RRBs as well as physical and sensory needs. It is perhaps unsurprising that these two areas are linked to the diagnostic criteria for autism in DSM-V and ICD-11 (APA, 2013; WHO, 2020) and relate to specific difficulties teachers may need to address with autistic pupils. As these areas have specific relevance to autism as shown by the link with the diagnostic criteria, these intrinsic barriers were separated from the second subtheme of Theme 1 which incorporated skill deficits resulting in barriers to learning.

Within this subtheme, a number of different areas linked to RRBs were considered. Rigidity in thought was mentioned and examples were given, such as a pupil not being able to use initiative in varying circumstances and only operating within particular parameters. The discussion also involved ritualistic and habitual repetitive behaviours which can interfere in learning. One participant gave the example of a pupil who needed to conduct certain rituals before they could access learning, saying, "*you couldn't get anything out of her until she had done that.*" Another participant stated, "*it's about managing that, saying it's ok for you to do that but not all of the time because that is restricting your experiences in school.*" Restricted behaviours and interests were mentioned as impacting upon pupils' abilities to learn, particularly for verbal pupils with more moderate intellectual disabilities who may find it difficult to move on from preferred activities and topics of conversation. Difficulty with change and transitions was mentioned on multiple occasions across both focus groups. One participant explained, "*it takes them a long time to then understand that change which can take a whole lesson and you then don't have the opportunities to get out of them what you wanted in that lesson.*"

Physical and sensory needs were also identified as potential barriers to learning. Both sensory seeking behaviours and sensory aversions were discussed as well as basic physical needs. Participants mentioned sensory distractions as well as the specific sensory requirements of some pupils before they could settle to learn. The diversity of sensory needs was recognised and one participant acknowledged that, "*their sensory regulation is on a huge spectrum isn't it?*" Basic physical needs such as hunger, sleep difficulties and medical needs

were also seen as possible barriers to learning, as they would for any pupil. It was identified that "*there seems to be some patterns… physiologically*" for autistic pupils, with bowel difficulties given as an example. Sensory input was considered across both focus groups to be helpful for pupils who may have sensory-seeking behaviour and it was viewed as necessary for teachers to be aware of the sensory needs of individual pupils and any such difficulties pupils may be facing.

5.3.1.2 Atypical Skill Development Resulting in Barriers to Learning

These areas were identified as skill gaps which would result in barriers to learning and included emotional regulation, learning behaviours and difficulties in functional communication. Deficits in these skill areas may result in barriers for any learner, however difficulties are likely to be exacerbated for pupils on the autism spectrum due to their specific needs.

Emotional states were discussed by many teachers throughout both focus groups as significantly impacting access to learning. Anxiety and frustration were specifically identified as problematic alongside difficulties in children recognising, regulating and coping with their emotional states. There were discussions of "*recognising their own emotions*" and supporting children to "*cope with the situations when they go wrong.*" Anxiety and heightened emotional states may be considered consequences of other difficulties rather than barriers in themselves, however it was clear from the focus group discussions that the regulation and managing of pupils' emotional states were one of the biggest challenges for both pupils and teachers.

Another idea that emerged from conversation involved learning behaviours. Areas discussed as making a positive difference to learning included attention skills, focus, readiness to learn, engagement, enjoyment in learning, confidence, self-esteem, risk-taking and problem solving. The discussion suggested that, although these skills are 'expected', they are not usually taught directly or assessed through the curriculum regardless of how necessary they are for learning to begin. One participant expressed the idea that "*learning isn't just about learning subjects.*" Another teacher gave an example of the impact of small steps in these learning behaviours which actually represent great progress for the individual pupil:

> "*We've had some children that won't even come into the class and then you look, say, six months later they're in the class… that is progress.*"

Communication was a final area discussed in relation to skill gaps resulting in barriers to learning. Almost all teachers across both focus groups strongly agreed that an effective form of communication, in whatever way was most suitable for that child, was necessary for learning to be able to take place. Teachers spoke of pupils being able to communicate their wants, needs and feelings and highlighted the need for the communication to be spontaneous and functional. One teacher gave the example of a pupil who used their own variations of signs to communicate which meant only staff who knew them well understood what they were trying to say. They explained, "*she tries to communicate and then gets really angry and then the behaviours start so there's a barrier already with her even thinking, 'oh I won't even bother then.*" The link between lack of functional communication abilities and other areas resulting in learning barriers, such as frustration and anxiety, was addressed throughout the discussion.

Emerging from the discussion was that the interactions between the characteristics or needs of autistic pupils and their potential skill gaps provide a profile of barriers to learning that autistic children in special schools may face. The interaction of these various difficulties along with the heterogeneity of autism results in barriers which appear specific to this group of pupils.

5.3.2 Theme 2: "Only when you are at a place where you understand that child can you really… help them to overcome their barriers" - Overcoming Barriers to Learning

This theme emerged throughout both focus groups as teachers discussed factors which most helped their pupils on the autism spectrum to overcome barriers to their learning. Three subthemes were evident; *pupil-teacher relationships, collaboration* and *links with parents/home.*

5.3.2.1 Pupil-Teacher Relationships

This subtheme was evident through the entire course of both discussions with group agreement on the importance of the relationship between a pupil and teacher. Teachers talked about really knowing their pupils and explained that "*we get to the point where we can read them well.*" Participants also talked about the relationship from the opposite perspective and highlighted the importance of the pupils knowing and trusting the teacher. The length of time it can take to build those teacher-pupil relationships was acknowledged as a difficulty, with

one teacher discussing the pupils who had been in their class for nine months and stating, "*I've only now really got to know them.*"

Further difficulties were mentioned, such as the fact that pupils may not engage with an unfamiliar staff member as they would with someone they know well. It was recognised that if there is a good relationship with teachers then the pupils are "*more inclined to show you what they're capable of.*" There was a strong accord among teachers that knowing a pupil well could make a huge difference to their progress in accessing their learning and overcoming the barriers identified in Theme 1.

5.3.2.2 Collaboration

Collaboration between teachers or school staff, both within a school and between schools, was described as beneficial to support pupils and effectively identify and assess pupil progress. Teachers commented, "*we do discuss it a lot as a class team,*" "*we can talk about that together, we can pull our ideas together*" and "*sharing our expertise is key.*" Collaboration was also mentioned as a way of reducing teacher workload. Discussing pupils' needs and progress with support staff, other teachers and between schools was considered useful as it helped creativity with teaching and intervention strategies as well as addressing the areas described in Theme 1.

5.3.2.3 Links with Parents/Home

Teachers across the two focus groups recognised the importance of building relationships with parents and having an understanding of pupils' home lives. One of the focus groups highlighted this as a crucial aspect of overcoming barriers. Teachers described progress that parents were particularly pleased with, which mainly involved socially appropriate life skills such as being able to take their child on holiday or out to a restaurant. When talking about these skills, participants described "*for some of our parents it must be some of the biggest pleasures*" and "*we've had some parents thank us profusely*." Teachers identified that more support can be provided for parents when there is an awareness of circumstances at home. There seemed to be genuine positivity from teachers when they talked about collaborating with parents, whether sharing positive news about the pupil or receiving input from parents. The benefit of consistent approaches across home and school was also mentioned when discussing pupil progress.

5.3.3 Theme 3: "That's a big part of what we need to teach them in school" - Priorities for Autistic Pupils

Theme 3 of the discussion included other skills that teachers thought it important for their pupils on the autism spectrum to make progress with. This theme mainly incorporated the application of skills and the two subthemes included *generalisation to the real world* and *community access and engagement*. Some ideas in this theme also related to supporting parents and improving life for the pupil and family at home, showing evidence of overlap with Theme 2.

5.3.3.1 Generalisation to the Real World

Teachers stressed the importance of their pupils being able to generalise skills. They stated that it wasn't enough to just teach skills in school but that pupils needed to be able to "*transfer those skills*" to "*different settings*" and display them with "*a variety of adults.*" The teachers repeatedly mentioned independence and self-care when discussing priorities for their pupils and highlighted specific skills such as hanging up their own coat and bag, using the toilet appropriately and using a knife and fork. Participants talked about teaching generalisation of skills in order for pupils to being able "*to cope in the world*" and preparing pupils for life after school. There were discussions about supporting pupils to be present around large groups of people and to communicate with others outside of the structured environment in which they were taught. It was also acknowledged, however, that teachers face limitations to teaching skills in meaningful contexts. Although pupils may have opportunities to access the community through school, most teaching is restricted to school settings and it can be difficult to fully prepare pupils for everything they may come across in the 'real world'. As one teacher identified, "*you can't replicate or anticipate every single situation or variation of situation that that person may come across. It's impossible.*"

5.3.3.2 Community Access and Engagement

Within the discussion of preparing pupils to manage outside of school, teachers talked about access to and engagement in the community. Accessing the community was discussed both in terms of the pupils being able to be physically present in public areas such as swimming pools and restaurants and also in terms of the public accepting the presence of the children in these community spaces. A large part of this subtheme involved prioritising the teaching of socially appropriate behaviours. Teachers mentioned the difficulties that may be

faced for pupils and their families if they display behaviours such as not being able to sit in a public place or removing clothes. Two participants discussed an example of one such important life skill:

> "*If they're going to the toilet… to not come out of it with your pants around your ankles, you know, so you pull your pants up, pull your trousers up [shut the door in the first place].*"

The teachers talked about pupils having more opportunities in the future if they could learn skills such as these. A link was also evident with behaviours discussed in Theme 1 such as managing emotions, functional communication and RRBs.

5.3.4 Theme 4: "We've assessed it this way but is that really mastery yet?" - 'True Mastery'

During discussions on assessment and important skills and progress, an idea seemed to filter through the conversations which has been termed 'true mastery'. Through the discussion, teachers questioned and attempted to define 'true mastery' and then considered if and how it might be assessed.

5.3.4.1 Rote Learning vs Real Understanding

Teachers identified that there is a difference between rote learning and real understanding. They acknowledged that, perhaps in part due to autistic pupils' restricted behaviours, skills are often learnt by rote and then performed within very specific parameters. Participants recognised that, when assessing skills, it was easy to assume that a pupil could do something with understanding when actually it was rote learnt. One teacher gave the following example: "*It's like a young child when they first learn to count… they can learn to count to five but they don't know what five is.*" The ability of teachers to distinguish between rote learning and real understanding was considered important, with one participant explaining, "*there's that danger… that we teach and they learn it in a rote fashion.*" Another teacher used an example of echolalia, saying "*you'd think they understand what you were saying but actually they're just repeating.*"

5.3.4.2 Do We Know What True Mastery is?

During both focus group discussions, teachers questioned how true mastery could be identified or whether it could even be defined. They questioned, for example, "*how do you*

really know when a child's completely generalised a skill?" Evaluating mastery was considered an ongoing process during assessment, with some participants believing that true mastery can never be defined. One teacher asked, "*[is it] five different environments that they show it? Is it ten? How do you say what is the ultimate generalisation of that skill?*" Some teachers attempted to resolve this question by suggesting that there is an element of 'just knowing' that a child has mastered a skill. It was also suggested that the photographic and video evidence 'proving' a pupil has mastered a skill isn't necessary to the extent that it is sometimes required by schools.

5.3.4.3 Interpretation can Vary

Participants brought up potential difficulties around variation in teachers' interpretations of skills, behaviours and abilities. They acknowledged that differing opinions or values can lead to varying perceptions of the same skill and they talked about the fact that "*different staff members might have different thresholds*" when assessing a skill. One participant also acknowledged that assessment can be reflective of a teacher's abilities as well as their opinions and values. A number of participants identified, with agreement from the group, that there is pressure on teachers to show progression in assessments, regardless of circumstances, "*because children have always got to make progress… and you're questioned if they don't.*" Potential resolutions to these difficulties were suggested, for example ensuring precision in the assessments themselves and the recording of the progress: "*it's being so precise that people know exactly what it is they've done.*" Comments were also made about the importance of uniformity in assessments, both within schools and across schools, to ensure there is a common language and understanding around the progress assessed.

5.3.4.4 Assessments Reflecting True Ability

Teachers talked around the idea that assessments don't always reflect the true ability of a pupil. They mentioned the importance of being able to evidence breadth of learning and record the different circumstances in which a child might be able to demonstrate a skill. This subtheme linked to the subtheme of *rote learning vs true understanding*, with the suggestion that assessments often don't allow differentiation between these types of learning. One participant explained it was sometimes difficult to determine skill level, saying "*it might not necessarily be an indication of… whether they can do that securely every time they're presented with that in different situations.*" When talking about the different needs of pupils

on the autism spectrum, another participant stated, to mumbles of agreement from the group, "*assessment tools don't reflect that do they? They're too rigid.*"

5.3.5 Theme 5: "The actual assessments don't tell the full story" - Assessing the Bigger Picture

In both focus group discussions, teachers spoke of the 'bigger picture' being particularly important for autistic pupils. There was discussion around assessments needing to take account of the context and situation when demonstrating skills and allowing for individual aspects of a pupil's learning.

5.3.5.1 Looking at Pupils Holistically

Teachers talked about pupils' learning happening alongside other learning in the wider context of their lives. Participants highlighted that a central part of teaching at school is about "*developing the whole child.*" One teacher spoke of the artificiality of considering skills in isolation from this wider context. They explained:

> "*I can pull in lots of observations that make a holistic overview of where that child is in many, many different areas because I just don't think you learn maths like that... I think it's a mixture of activities in a context that's appropriate to them and the situation.*"

This holistic approach of looking at the whole pupil rather than looking at individual skills in isolation appeared important to teachers, particularly when it came to assessment.

5.3.5.2 Personalisation/Individualisation

Teachers recognised the heterogeneity of autism, mentioning that autism is a spectrum condition and that each child is "*unique and individual*" with a diverse range of needs. They discussed that they often need to adapt assessments to suit the needs of the pupils they teach. One example was the necessity to modify statutory speaking assessments for pupils who were pre- or minimally verbal and it was noted that the onus was on the teacher to adapt assessments accordingly. Some participants spoke of a desire for "*bespoke assessment systems*" which allow an element of personalisation. One participant was sceptical of this possibility, however, and stated, "*a perfect catch all cover all assessment system… I can't imagine it, not something that could cover every single child.*" Another participant responded by suggesting that it is up to the teacher to select appropriate assessments and commented, "*I*

think it's our responsibility maybe to find other tools for certain children to make it more individual because, as you say, not one size fits all." One participant, however, had concerns about this in terms of teacher workload and stated reservations about teachers "*reinventing the wheel*" rather than collaborating and sharing effective assessment practice within and across schools. This subtheme captured the idea that, if it could be done in a practical sense, personalisation of assessments and the option to individualise them for individual pupils would be useful.

5.3.5.3 Importance of Recording Nuances and Subtleties in Behaviour

Teachers were categorical in talking about how very small steps of progress could be important for a pupil and their family. When talking about skills and behaviours identified in Themes 1 and 3, seemingly small progress was described as "*so important,*" "*really small yet massive in their achievement*" and "*such a big achievement and step for him.*" Teachers mentioned how academic skills are often prioritized, even though the progress which may seem small is often a bigger achievement and more important for that pupil. An example was given of a pupil who had managed to communicate how they felt more appropriately, even though they still had difficulties in their classroom learning. Teachers expressed how, in spite of the importance, assessments didn't allow for the recording of such small amounts of progress.

5.3.5.4 Regression

The conversations around the concept of regression developed in an interesting way. In one focus group, initially there was a reluctance to consider regression in skills, abilities or behaviour as something that should be recorded or assessed. Teachers asserted "*no, we don't do it,*" "*absolutely no way we would ever be allowed to say that a child regressed*" and "*it's just so negative, isn't it?*" Gradually, the conversation altered after one of the participants questioned this established need to always make progress. They recognised:

> "*If they've reached a certain point then the expectation is that they go to the next level. But realistically you know that they've actually gone backwards… that needs to be noted.*"

At this point, the participants began to identify that there was nowhere for teachers to show regression in current assessments and, in reality, there are "*always times where their learning and abilities would regress.*" Both focus groups mentioned the summer holidays as times

when pupils might present with some regression in skills and behaviour. Participants also acknowledged that major changes at school or home could result in regression and they recognised the importance of identifying regression in terms of addressing the bigger picture for these pupils. One example was the extra support that was able to be provided for a family during a parent's illness which had only become apparent to the school as a result of the child's regression. The conversation also mentioned the difficulties of assessing regression, with teachers addressing the fact that once progress was recorded it couldn't be 'undone' on many assessment systems. Some teachers said they felt confident in addressing regression through Individual Education Plans and targets. However, participants stated that few assessments that were used had a facility to show regression. The groups agreed that where regression existed, it is important to be able to show it. During this discussion, one teacher noted the necessity of distinguishing regression from lack of generalisation in skills, for example if a pupil has a change of class teacher. This links to the conversation around difficulties these pupils often have with generalising and the challenge in determining at which point skills have been mastered.

5.3.6 Theme 6: "There's no way of really recording or measuring that impact that we know is huge" - Practicalities of Assessment

This final theme pervaded the discussion on assessment. It incorporated some of the practicalities of assessment that were brought up by teachers including the fact that there is nowhere to record certain progress. It also encompassed the fact that spiky profiles and non-linear progress made by pupils on the autism spectrum mean they often do not fit into the usual assessment frameworks or systems. As part of this theme, teachers also identified some aspects of assessment that are helpful and useful.

5.3.6.1 Nowhere to Record That

Throughout both discussions, participants on multiple occasions brought up the fact that there was nowhere to record some of the progress they were recognising. This was often in relation to the areas identified in the discussions on barriers to learning and teaching priorities. Teachers often expressed this in a questioning way, for example, "*where do I put this? Where do I write this down?*" and "*there's nowhere to say that so how do we put that down there?*" They mentioned that assessments often don't cover those small aspects of progress which were identified as important to teachers, pupils and parents. One teacher spoke of the creativity needed to "*try to make it fit*" in situations where progress can't be

recorded within an assessment system in a straightforward manner. As well as the practicality of having no way to record certain progress, teachers also suggested that there is some progress which *can't* be measured. One participant questioned, "*how do you showcase how willing someone is to want to learn?*" This was similar to the way teachers questioned whether true mastery can ever be determined.

5.3.6.2 Autism Not Fitting Assessment Frameworks and Systems

Teachers spoke specifically about the needs of their autistic pupils and the fact that they often didn't fit into the assessment frameworks and systems used. Teachers mentioned the "*spiky profile,*" the "*peaks and troughs*" that are characteristic of autism, and the fact that these pupils are "*not linear with their progress.*" One teacher commented:

> "*If you're going to look at national and statutory requirements for assessment it has always been the kind of cohort with autism that have probably least best fit the P levels.*"

The same teacher described children on the autism spectrum with intellectual disabilities as being the "*worst done by*" when it comes to assessment. Teachers also talked about the gaps in learning for these pupils and the need to revisit learning in order to maintain skills. One teacher stated that "*showing the pattern*" would be better than forcing the learning profiles of pupils on the autism spectrum to fit into a linear assessment framework.

5.3.6.3 Helpful Aspects of Assessment

Finally, teachers identified some aspects of assessment they found helpful and which they would like to see in future assessments. Teachers explained that it was useful when assessments signposted next steps in learning and assisted teachers to set pupil targets. Teachers recognised that this was particularly helpful in skills that autistic pupils struggle with learning. Teachers spoke positively of assessments which break down progress into small steps and help them to understand the progression of a child's skills in a particular area. One teacher mentioned identifying areas where there are interruptions to development or gaps in learning so that teachers know where further support or interventions are required. Early years teachers across both focus groups showed more satisfaction with the assessments they used and the way they assessed their pupils than the teachers who taught primary or secondary aged pupils.

5.4 Discussion

The themes developed from the data showed that special needs teachers consider important progress for pupils on the autism spectrum to be related to barriers to learning for these pupils. RRBs, sensory needs, functional communication, recognition and regulation of emotions and learning behaviours are likely to impact upon further academic or pre-academic progress made in schools. Therefore, it is no surprise that they are considered priorities due to their potential impact upon education. These areas relate to those considered in Chapter 1 and are also similar to those which teachers identified as difficulties in the study by Helps et al. (1999), supporting the idea that these areas may affect an autistic child's ability to learn and make progress in special schools. Interestingly, the teacher participants in this study did not include or describe 'behaviour that challenges' as a barrier to learning. Instead they closely considered behaviours or skill gaps which may *result* in behaviour that challenges and, in turn, affect access to learning.

Teachers also identified generalisation and application of skills to the 'real world' as priorities. This related to the importance of supporting parents and the home life of pupils and their families. Teachers spoke about the limitations for parents when a pupil's behaviour is not considered to be 'socially acceptable' and the opportunity to address these areas is likely to be directly affected by the barriers to learning that were identified. Although the barriers to learning which teachers identified were discussed in the context of a school environment, it is possible, and perhaps even likely, that addressing those barriers to learning would also positively impact upon pupils' home life and improve access to and engagement in the community for the pupils. The discussion around collaboration and relationships between teachers and parents confirmed previous research findings that these partnerships are "best practice" in the education of children on the autism spectrum, positively impacting upon outcomes and care (e.g. Azad and Mandell, 2016; Syriopoulou-Delli et al., 2016, p.2). The barriers and priorities that teachers discussed can be linked to various quality of life and family quality of life domains (Schalock, 2004; Hoffman et al., 2006).

When considering these areas in terms of assessment, teachers spoke of having skills and behaviours broken down into small steps and for assessments to recognise particularly small aspects of progress. Teachers identified that the "*nuances and subtleties in behaviour*" can be of particular importance in terms of pupil progress and, therefore, are the types of progress that need to be recorded.

A number of other aspects of assessment discussed by teachers need consideration due to the potential conflict between them. Two such dichotomous areas involved the discussion around precision in assessments to minimise the variation in interpretation by different teachers contrasted with the conversation on the advantages of potential personalisation of assessments. Precise assessment items help to preclude varying interpretations by teachers when assessing pupils and this relates to the reliability of the assessment. Inter-rater reliability may therefore be important in order to ensure that different teachers or raters are assessing the same skill or behaviour. However, precision may result in the assessment being what teachers described as "*too rigid*" and may require teachers to adapt the items themselves in order to fit the needs of their pupils. Personalisation may affect both the reliability of the assessment if different raters are adapting or individualising an assessment separately and the validity of the assessment if the assessment item has been personalised to a point that it no longer measures what the original item intended to measure. Linked to personalisation, teachers also spoke about the need to have a holistic overview of a pupil in order to assess their progress. They recognised that when a teacher knows a pupil well, they are more able to consider depth of learning. This supports the use of report measures in schools, where teachers assess skills based on their knowledge of the child. Good knowledge of the pupils' abilities, communication and motivation may allow a teacher to overcome some of the difficulties in assessing autistic pupils with intellectual disabilities which were considered in Chapter 2. However, the need to know a pupil well may, again, affect reliability of an assessment. In light of the teachers' discussion, if an element of personalisation can be added to a teacher report measure while still ensuring reliability and validity, the assessment could prove particularly helpful and useful for teachers. These discussions provide practical examples of assessment development and evaluation concepts as they may apply to educational assessment, as discussed in Chapter 2.

Two further ideas which need to be reconciled included the need for assessments to address next steps which contrasts with the desire for an assessment to be non-linear. Linear assessments, which track development, inherently show teachers what is next in the acquisition or development of a skill. Teachers acknowledged during the discussions that, due to the spiky profiles and unusual learning patterns of autistic children in special schools, learning and progress may be scattered and therefore difficult for a linear assessment tool to capture. This, again, links to autism-related characteristics such as the uneven cognitive profile discussed in Chapter 1, and the resulting difficulties in assessing these pupils. It may

be beneficial to diverge from traditional linear conceptualisations of skill development when creating assessments for these pupils and explore ways which non-linear progress can be taken into account while still providing next steps.

Finally, teachers discussed the need for assessments to acknowledge and account for regression in skills. Teachers identified that assessments often don't allow for the recording or reporting of regression and may need to be adapted to ensure current and future progress reports are accurate. They recognised that regression may be part of the bigger picture for some pupils and can also potentially alert them to other needs that a pupil may have.

5.5 Limitations

This research has given special needs teachers the opportunity to outline their views on barriers to learning, important progress, teaching priorities and assessment of autistic pupils in special schools. There are, however, a number of limitations of this study. Firstly, a small sample was used from two schools in one area of England. The discussions and themes were similar across the two schools and it is therefore encouraged that the research is replicated with special needs teachers in other areas to determine if these views are representative of those in other special schools. Secondly, the focus of this research is purposely narrow; the participants were teachers from special schools and the discussions were focused upon autism. Caution should be taken if results are to be interpreted outside of this context. A further limitation of focus group discussions relates to the facilitation by the researcher. Flick et al. (2004) identified the importance of the focus group facilitator creating the right conditions conducive to the particular discussion. They proposed eight principles to follow when conducting focus groups including addressing the entire group rather than individuals, suggesting instead of prescribing topics and appropriate questioning. As far as possible, these principles were adhered to during the focus group process. Finally, it may be that issues discussed were misinterpreted or lost meaning during the qualitative analysis. The possibility of this was mitigated through the use of participant checking, a reflective journal kept during the research and discussion with the doctoral supervisors throughout the process.

5.6 Conclusions and Recommendations

The focus groups provided detailed information about the areas that the special needs teachers considered to be barriers to learning for their autistic pupils. Teaching priorities for this group of pupils were also highlighted along with aspects of educational assessment that

are valued when assessing these pupils in special schools. The focus group data allowed for a more in-depth understanding of areas which impact upon these pupils' learning and, in turn, supported the definition of the 'barriers to learning' assessment construct. Teachers also provided specific examples when talking about skill areas which were then useful when assessment items were drafted. Various aspects of the focus group discussions were informative in terms of the practical features of the new assessment. The format of the assessment took account of the ideas discussed above, notably potential elements of personalisation, accounting for regression and the non-linear progress often shown by autistic pupils. Aspects of reliability and validity were considered when constructing and evaluating the assessment, for example the possibility of varying teacher interpretations when assessing pupils. The construct, related skills and features of the assessment will be outlined further in the description of the ABLE-Autism in Chapter 7.

This research also has some broader implications for the development of educational assessments. Most special needs teachers have very good knowledge of their individual pupils' needs and priorities for their progress. In spite of this, teachers are not often given a voice when it comes to the way that progress is identified and recorded. In order to ascertain the face and content validity of assessment tools while also ensuring that assessments are useful and valued by teachers, it is imperative that teachers have input during the assessment development stages and that any concerns and preferences about the assessment content and method are taken into account. It is recommended that teachers are involved in all aspects of the assessment development process, particularly prior to the development of an assessment when defining parameters and considering the purposes for which the assessment will be used. Teachers should have input into the items to be used within assessments, and their judgements about the utility of an assessment should be sought and considered. Unhelpful and ineffective assessments can measure irrelevant skills, add to teacher workload and waste valuable teaching and learning time. It is necessary, therefore, for assessments to be developed for the unique needs of pupils on the autism spectrum, for the specific context of use in special schools and with input from the individuals who use them – the teachers. Therefore, as well as suggesting content and features for the assessment, the results from the focus groups also provided further justification for the need for a robust and effective assessment developed specifically for the needs of autistic pupils with intellectual disabilities which can be used by teachers in special schools to assess barriers to learning.

Chapter 6. Study 2b – Delphi Exercise

6.1 Introduction

The construct of 'barriers to learning' for pupils on the autism spectrum in special schools was further defined following the focus groups. The final definition and parameters of the construct will be detailed in the description of the ABLE-Autism in Chapter 7. Based on the broad areas and specific examples provided by teachers in the focus groups as well as the review of the literature, a list of skills and behaviours was created. Further information about the creation of items will be provided in the detailed description of the assessment in Chapter 7. A group of special educational needs teachers were then invited to take part in a modified two-round online Delphi exercise in order to examine the set of proposed assessment items.

As discussed in Chapter 2 and recognised in the systematic review, when new assessment tools are being developed, it is important that stakeholders are provided with an opportunity to comment on the assessment content to ensure face and content validity (Boateng et al., 2018). Content validity is arguably one of the most important measurement properties of an assessment (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). The assessment items must be relevant, comprehensible and comprehensive in respect of both the domain or construct to be measured and the population in question (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). Content validity is usually evaluated by experts in the relevant areas. At this stage in the assessment development it was decided that special needs teachers would be consulted due to their wider training, knowledge and experience in assessment, curriculum, specific interventions and pedagogy. As the professionals who will be regularly using, facilitating and interpreting the assessment, they were considered the starting point for the initial selection of items. As mentioned in Chapter 1 and supported by the findings in Chapter 5, key skills and behaviours for pupils on the autism spectrum in special schools often relate to future outcomes in adulthood and quality of life outcomes. Therefore, it is appropriate for other stakeholders such as parents and pupils to provide input when a new assessment is being developed and this will be a recommendation for further evaluation of the ABLE-Autism.

It has been argued that the value of focus groups is limited to generating ideas and hypotheses (Flick et al., 2004). For this reason, the Delphi exercise followed the focus groups

to further consider the content of the assessment. As described in Chapter 3, a Delphi exercise is an established method of determining consensus on the content of an assessment by a group of experts (Boateng et al., 2018). The quantitative nature of the Delphi exercise complemented the qualitative data produced by the focus groups and allowed for clear decision-making about the rejection or retention of items. The opportunity for participants to individually and anonymously have their say during the Delphi exercise also countered any potential influence of group hierarchy or influence during the focus groups.

The items derived from the focus groups, review of the literature and systematic review were grouped into five categories. These categories were labelled as Learning Behaviour Barriers, Restricted and Repetitive Behaviour Barriers, Emotion and Behaviour Regulation Barriers, Sensory Barriers and Functional Communication Barriers. The aims of this research were to consider:

(a) the relevance of a list of potential assessment items,
(b) the comprehensibility of those items and
(c) the comprehensiveness of the list of items by addressing whether any skills or behaviours were missing from the list.

Amendments to items were made where necessary and each item was either accepted for inclusion in the assessment or rejected. The three areas of comprehensibility, comprehensiveness and relevance were selected to ensure sufficient content validity as suggested by the COSMIN manual (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018).

6.2 Methods

6.2.1 Participants

Purposive sampling was used to recruit teachers to participate in the online Delphi exercise. Thirteen of the focus group participants who had expressed an interest in further participation were invited by email to participate in the Delphi study. Teachers in the author's professional networks were also emailed and invited to participate. Participant expertise is important within a Delphi exercise and, dependent upon the purpose, different stakeholders can be regarded as 'experts'. For example, patients rather than clinicians have been used in health-related Delphi studies and are considered "experts by experience" (Law and Morrison, 2014, p.1348). The eligibility criteria for the current study were: (a) that the participants were

qualified teachers, (b) currently working in a special school and (c) had recent experience of working with pupils on the autism spectrum. As a school assessment is likely to be used by all teachers regardless of experience, expertise or specialism, it was felt that criteria for further selection (e.g. for a teacher to have been teaching for a specific amount of time) would potentially affect the results and, therefore, no further criteria were required. Eleven participants took part in round one and nine in round two. There is no general consensus for the number of participants in a Delphi exercise. Minimum suggested numbers are usually around 7-10 (Day and Bobeva, 2005) but Delphi exercises have been conducted with as few as three participants (Skulmoski et al., 2007).

6.2.2 Ethical Compliance

The information provided to focus group participants also included information about the Delphi exercise. A second information sheet and a second consent form were provided in an electronic format using Qualtrics and this was completed before the Delphi questionnaire was made available. The information sheet and electronic consent form was also provided for the second round of the Delphi if participants wished to revise it. Informed consent was obtained electronically through an online version of the consent form and by completing and submitting the questionnaire.

6.2.3 Procedure

6.2.3.1 Modifications

There were a number of differences and modifications made to the established Delphi method for the purposes of this research. These modifications included a reduced number of rounds and a change of response in the second-round questions.

Although a three-round Delphi exercise is considered typical, two rounds or even one round may be sufficient depending upon the research (Skulmoski et al., 2007). Although the first round of a Delphi exercise is often qualitative, extensive qualitative data had already been collected from the two focus groups. It was therefore felt that the Delphi exercise could begin with more of a focus on directly rating the relevance, comprehensibility and comprehensiveness of potential items. In addition, a two-round Delphi reduced burden on teacher participants in requiring them to rate and comment upon the same items numerous times (Walker et al., 2019). Further, with a dropout rate of 18% in round two, and as a high number of items often results in a significantly lower response rate (Gargon et al., 2019),

there was a risk of compromising the quality of the data and subsequent results through a particularly low response rate in further rounds. Delphi exercises may continue until a consensus is reached and, with over 80% of the original items meeting the high threshold for consensus in order to be retained in their original or amended format after two rounds, it was determined that a third round was unlikely to provide any significant change in the overall consensus for remaining items. The two-round Delphi, therefore, was sufficient in order to answer the research questions and establish consensus on the items to be included in the ABLE-Autism.

A second modification to the Delphi procedure was the decision to amend the response options to 'yes' or 'no' for the second round. A traditional Delphi exercise provides participants in subsequent rounds with the same questions in the same format and asks them to reconsider previous responses in light of feedback from preceding rounds. However, the response options for the second-round questions were amended in this modified Delphi for the following reason. The spread of scores were low in round one with responses on only eight of the 86 items having a standard deviation (SD) ≥ 1 for comprehensibility and 11 items with a SD ≥ 1 for relevance, with the highest SD of any item being 1.3. These first-round results included more positive and less spread responses than originally expected. In order to differentiate between items which had a strong or moderate consensus on inclusion, a high threshold was decided upon when defining consensus. The first-round responses showed a strong consensus for the inclusion of certain items which were retained after round one and therefore not included in round two. However, the results did not show a similar obvious consensus for exclusion (e.g. all but one item which did not reach the high threshold set for inclusion still received over 50% of the responses rating them as relevant and comprehensive). As the decision to include or exclude items was ultimately dichotomous and the nuance of a scale did not provide strong evidence of either inclusion or exclusion, the decision for participants to clearly indicate inclusion or exclusion using a dichotomous response was taken. The decision to change the response options in round two was also influenced by teacher workload, as participant burden and fatigue were considered if participants were asked to complete the same questions on the same scale a second time.

6.2.3.2 Round One

Following a pilot test, round one of the modified Delphi exercise contained 86 items and asked participants to rate each item on comprehensibility (how well they understood the

item) and relevance (how important they thought the item was to assess) specifically in relation to pupils on the autism spectrum. Items were rated for each question on a four-point Likert scale ranging from 1 (i.e. very easy to understand or very important) to 4 (i.e. not at all easy to understand or not at all important). As the end result is the dichotomous inclusion or omission of an item in the final assessment, a mid-point or neutral value was not used in order to avoid central tendency bias, social desirability bias and the distortion of data (Nadler et al., 2015). At the end of each section an open-ended question on comprehensiveness asked if any skills or behaviours in this area were missed or should be added to each of the five categories. A final open-ended question on comprehensiveness at the end of the questionnaire asked for any further comments about the items or assessment as a whole. An optional comment box was also provided for each item where participants could add comments if they wished. As a result of the earlier focus group discussion, items were worded positively (i.e. describing what the pupil can do rather than what they cannot) and potential personalisation of the assessment was outlined by having the items refer to a fictional child, 'Charlie'. The first round of the questionnaire was left open over a period of six weeks as its distribution coincided with the school summer holidays. This allowed teachers who may not be available as they would be in term time an adequate amount of time to respond. Participants were twice informed by email of the extended time allowed to complete the questionnaire. Eleven responses were received in total. In the first round, 4 questions had one missing response each.

6.2.3.3 Round Two

The participants who had completed and provided consent in the first round were given access to the second-round questionnaire. The consent form and participant information sheet were available if participants wished to review them. Participants were provided with feedback on the first-round responses and informed that the items included in the second-round questionnaire were the items which less than 80% of teachers rated as either able to be understood or important to assess. The second-round questionnaire contained 27 items which had been amended after feedback from the first round. The second round featured 'yes' or 'no' responses to the two questions on comprehensibility and relevance. As a number of items had already been retained and were therefore not included in the second-round questionnaire, the question on comprehensiveness was removed so that participants did not request skills and behaviours which had already been retained with the intention of being included in the assessment. Again, an optional comment box was provided for each item as

well as a final, open-ended question at the end of the questionnaire where participants could opt to make any further comments. The second-round questionnaire was open for two weeks and participants were reminded by email when the questionnaire was approaching the date of closure. Nine responses were received with a second-round response rate of 82%.

6.2.4 Data Analysis

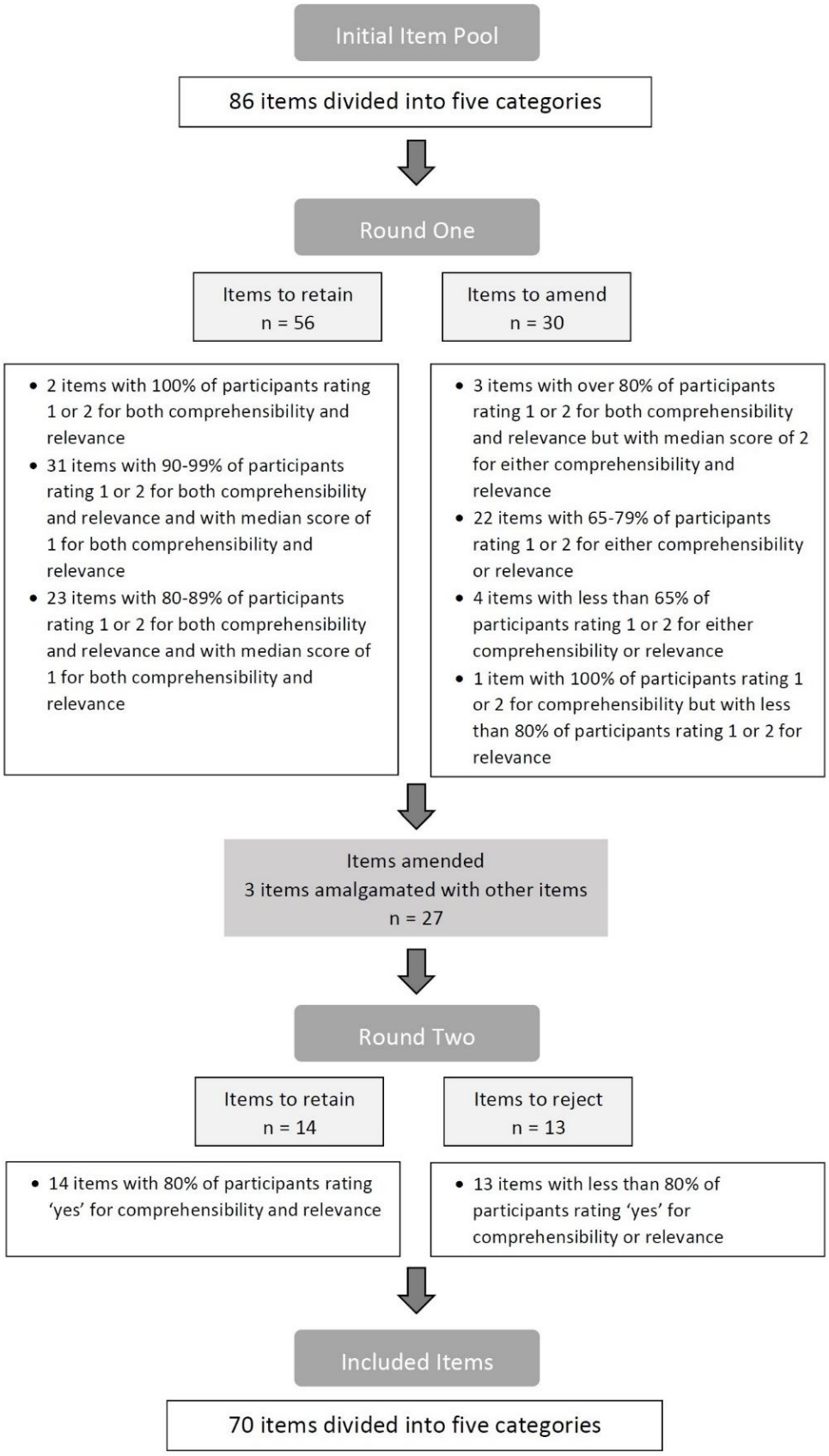
The criteria used in the analysis of data collected through the Delphi method may vary. Hsu and Sandford (2007) recognised that for a four-point Likert scale, at least 70% of respondents needed to rate on one side of the scale with a median of at least 3.25 (or 1.75 or lower for reverse scoring). They identified that a median is an appropriate score to consider in a Delphi exercise as it shows converging opinion. As the current modified Delphi exercise considered comprehensibility, a high threshold for consensus was considered necessary and it was decided that an 80% consensus was deemed appropriate for item retention in this study (McConachie et al., 2018). The criteria used for consensus in round one was that: (a) 80% of participants rated the item as 1 or 2 for comprehensibility (very easy or quite easy to understand), (b) 80% of participants rated the item 1 or 2 for relevance (very important or quite important to assess) and (c) the median score for both questions was 1 (the top response). If the item met all three criteria then consensus was determined and the item was retained in its current form. Any items not reaching these criteria were eligible for amendment and then returned to the participants in round two.

A similar threshold for consensus was used in round two; that 80% or more of participants identified that (a) ‘yes’ they understood the item, and (b) ‘yes’ it is important to assess progress in the skill or behaviour.

6.3 Results

A summary of the findings from both Round 1 and 2 is outlined in Figure 2.

Figure 2. *Flow Chart of Results from Round 1 and 2 of the Delphi Exercise*



6.3.1 First Round Results

6.3.1.1 Retained Items

Eighty-six items were presented in round one, and a consensus was reached on 56 items as per the criteria specified above; these items were retained without any changes (Appendix 5.3). For two of these items, 100% of participants answered 1 or 2 to both comprehensibility and relevance. A further 31 retained items had over 90% of participants answer both 1 and 2 to both questions (agreed that the item was relevant and comprehensible) with a median score of 1. The additional 23 retained items had over 80% of participants answer both 1 and 2 to both questions with a median score of 1.

6.3.1.2 Items for Amendment

In total, 30 items were considered for inclusion in round two. Twenty-six items had less than 80% of participants rate either comprehensibility or relevance as 1 or 2, therefore not meeting criteria (a) or (b) for inclusion after round one. Three additional items had over 80% of participants rate 1 or 2 for both questions but, with a median score of 2, did not meet criterion (c). All of these 29 items were set aside for amendment before being taken forward to the next round. One further item was rated as able to be understood by 100% of participants (1 or 2 for comprehensibility) but only 75% of participants rated 1 or 2 for relevance meaning that it did not reach criterion (b). As it was considered comprehensible by all participants, this item was not amended but put back into the round two questionnaire to determine whether a consensus could be reached.

6.3.1.3 Additional Comments

Eight of the 11 participants made at least one comment in the optional comment boxes in round one. Three teachers suggested additional areas to include. Skills such as "*accepting a change or when things don't go according to plan*" were identified by participants in the comment box at the end of early sections, however these skills were covered in later sections. Some of the items which were taken forward to the second-round questionnaire were amended to clarify or to include suggestions from teachers in round one. For example, the suggestion of working with peers in different roles was incorporated when the item *Charlie will sit for the duration of a group activity* was amended to *Charlie will appropriately engage in a group activity with peers*.

Further comments on specific items often addressed the scope of the item and asked for more information such as prompt levels. As an aspect of personalisation is intended to be incorporated in the teacher responses in the final version of the assessment, items were not amended to be more specific at this stage and this will be considered further in the discussion section of this chapter. The intentions behind some comments were not entirely clear in terms of the survey questions, e.g. one comment stated that "*this is very difficult to teach non vocal children.*" It was unclear how to act upon these comments but, where possible, they were taken into consideration when amending the items for round two.

6.3.2 Second Round Results

Items which did not meet the criteria for retention after round one had their wording amended for inclusion in round two. It was determined that three of the items could be amalgamated with other items after rewording. For example, the item *Charlie will show awareness of different expectations in different rooms/settings* and the item *Charlie will show awareness of different behaviours being appropriate/inappropriate depending on the setting* were joined to form the amended item *Charlie will adjust their behaviour depending upon the expectations of the environment or setting*. In total, 27 items were presented in the round two questionnaire.

6.3.2.1 Retained Items

Responses from the second round indicated that 25 items were understood by 80% or more of participants. Fourteen of these items were also considered to be important to assess by 80% or more of participants and, therefore, were retained in their amended version for the final version of the assessment (Appendix 5.4). This led to a total of 70 items to be included within the ABLE-Autism.

6.3.2.2 Additional Comments

Five participants made optional comments which covered varying interpretations, positive feedback on the items and stipulations on items (e.g. "*yes if meaningful*"). One comment identified that "*all assessment items are well thought out and easily explained.*" As in the first round, a number of comments related to the specific level of support or interpretation of words used within the item. For example, one comment in the final optional comments box suggested there may be variations in interpretations around the word "*appropriate*". Again, this was addressed through aspects of personalisation in the final

version of the assessment but is an issue, as recognised in the focus group study, that will be considered further in Chapters 7 and 9.

6.4 Discussion

The modified Delphi exercise identified a total of 70 items which were included in the ABLE-Autism. Teachers rated the original list of items, devised from the literature and teacher focus groups, for comprehensibility, relevance and comprehensiveness as suggested by the COSMIN standards (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). Showing a consensus in these areas provides preliminary evidence for face and content validity of an assessment (Boateng et al., 2018). Sixty-five percent of items were retained without change in the first round of the modified Delphi exercise and 52% of items presented in the second round were retained. The final list of items selected for inclusion from this modified Delphi study consisted of 14 Learning Behaviour Barrier items, 12 Restricted and Repetitive Behaviour Barrier items, 16 Emotion and Behaviour Regulation Barrier items, 10 Sensory Barrier items and 18 Functional Communication Barrier items.

The percentage of items retained was reasonably consistent across categories. Sixty-seven percent (14 of 21) of the Learning Behaviour Barrier items were included in the final assessment after both rounds of the modified Delphi exercise, with 77% of items (10 of 13) included in the Sensory Barriers category, 84% of items (16 of 19) included in the Emotion and Behaviour Regulation Barriers category and 86% of items (12 of 14) included in the Restricted and Repetitive Behaviour Barriers category. Only one item of 19 in the Functional Communication Barriers category did not reach the threshold for inclusion in the first round or in its amended form in the second round (95%). The fact that no single category contained a high percentage of rejected items suggested that the five categories reflect areas which teachers believe are important domains when assessing pupils on the autism spectrum within special schools. The Functional Communication Barrier category contained the largest number of retained items which suggested that teachers rate a variety of functional communication skills as important to assess when considering and reporting pupil progress in schools.

Thirteen items were not considered relevant for assessment by a sufficient percentage of teachers. Items relating to showing pride or enjoyment in completing an activity and tolerating sensory input were rated among the least important skills or behaviours to assess despite these areas being brought up in teachers' focus group discussions as important areas

of progress. It may be that, as discussed in Chapters 2 and 5, the practicalities of translating these skills or behaviours into items for teachers to use to assess pupils means that the skills are difficult to capture. Similarly, the comments by teachers on possible interpretations of words included in the assessment also reflects teachers' concerns expressed in the focus group study around ensuring that descriptions are precise to avoid variations in the understanding of assessment items. The balance between assessments being specific enough to ensure reliability yet flexible to allow for the ways different pupils may display a particular behaviour or skill is difficult to strike when considering the wording and content of assessment items.

Six of the items retained in round one received one score of 4 for comprehensibility (not at all easy to understand). Although these items were included as they exceeded the 80% threshold for both questions with a median of 1, they were specifically marked to ensure they were considered further for comprehensibility in the evaluation study and this will be discussed in Chapter 8.

There are a number of practical implications of the Delphi study. The main implication is the selection of items for the ABLE-Autism. In line with the findings and recommendations from the systematic review and focus group study, teachers had a large role in the development and evaluation of the assessment. Allowing teachers to comment on potential items, suggest amendments, consider further inclusions and make decisions on the relevance of items, makes it likely that the items will show the areas of progress in the ways that teachers will find useful. A second implication of the Delphi study is that it outlines a method for teachers to give valuable input into the assessment development process through use of modified Delphi procedures and this will support the content validity of the assessment in future evaluations. It is likely that an assessment will be more useful and valuable if it was developed for the needs of the specific group who will be assessed, if it addresses the relevant areas and was developed and is evaluated in consultation with those who will use it.

6.5 Limitations

This modified Delphi exercise allowed for the selection of a number of different skills and behaviours to be included in the ABLE-Autism. There are, however, some limitations of the current study.

The sample was small and, although Delphi exercises can be conducted with fewer participants (Day and Bobeva, 2005), a larger population of teachers would have been preferable. A dropout rate of 18% from round one to round two is likely to have had a greater effect on the results when the initial sample is small, although second-round responses appeared broadly in line with the first-round responses and the results between rounds did not seem to alter in an unexpected way. The small number of participants choosing to participate in the modified Delphi exercise may be due to a number of factors including teacher workload, the initial length of the questionnaire and the fact that the availability of the questionnaire coincided with part of the school holidays. It is important to note that, although a Delphi exercise is considered a way for expert opinion to reach consensus, the responses from the group of participants may not be generalisable to the entire population being considered (Skulmoski et al., 2007). It is necessary to consider that the results may not entirely reflect the views of all special needs teachers and, therefore, further evaluation of the items in the finalised versions of the assessment is recommended with a larger sample. Similarly, this study considered teachers to be "experts by experience" (Law and Morrison, 2014, p.1348). The use of participants with specified experience relating to years teaching, level of responsibilities or teaching achievements was considered to conflict with the fact that the assessment is intended to be useful to all special needs teachers, regardless of experience. Therefore, participants' 'expertise' in the relevant area was determined by the fact that they were qualified teachers currently working in a special school and had recent experience of working with autistic pupils. This is also a limitation as, although this was not believed to be directly relevant for participant selection in the current study, it may be that newly qualified teachers have a different interpretation of important skills to assess compared to more experienced teachers. As suggested by Hasson and Keeney (2011), Delphi studies should not be taken as "indisputable fact" and instead be accepted as the opinions of a particular group at a particular time (p.1701).

Another limitation of the study is that, although other stakeholders were considered for participation, they were not consulted at this stage. Consulting families, other professionals and the pupils themselves where possible would be a useful way to triangulate the findings in the early stages of the development of this assessment. It would also ensure that the assessment content adequately covers important areas linked to quality of life, good outcomes in adulthood and future life opportunities for the pupils and their families. Evaluation and validation of an assessment is an ongoing process and, therefore, consulting

and including these stakeholders in assessment evaluation will be recommended for future research.

A further limitation of a Delphi exercise, particularly a mini-Delphi consisting of only two rounds, is the fact that misunderstandings may arise (Pan et al., 1996). Participants were able to contact the researcher by email or phone at any time if they had questions, though none did so. Certain comments were not able to be acted upon as they were not clear and the responses could not be followed up and clarified due to the anonymity of participants, although these were often from only one participant and their ratings of the items supporting their opinion were taken into account within the data analysis.

6.6 Conclusions

This chapter has outlined the use of a modified Delphi method to identify a total of 70 items endorsed by teachers which will be included in the ABLE-Autism. The use of the modified Delphi method to select the appropriate items to assess barriers to learning for pupils with autism in special schools provided initial face and content validity for the assessment. An evaluation of these items and the final version of the assessment as a whole using a relevant population in a school setting with teacher respondents followed the Delphi study in order to ascertain the measurement properties of the ABLE-Autism. As well as evaluation of the reliability and validity of the ABLE-Autism, it was necessary to continue to check content validity with teachers once the assessment has been used to ensure that the assessment is useful in practice as well as in theory. These further evaluations will be detailed in Chapter 8.

Chapter 7. The ABLE-Autism

Following the review of the literature, systematic review, focus groups and Delphi study, the initial version of the ABLE-Autism was developed. This chapter will describe the ABLE-Autism assessment and explain the reasons behind a number of decisions related to the assessment development and features. The process of the construct definition will be outlined and the purpose of the assessment will be explained. The assessment will then be described in detail with justifications provided for the format and features.

7.1 Construct Definition

'Barriers to learning' is a term often used in education, as discussed in Chapter 1. Its broader meaning in a school or educational context refers to anything which may result in difficulties accessing learning opportunities. In a general sense, this may include factors such as whether a pupil has SEN or the socio-economic status of the family. The construct was further defined in relation to barriers to learning relevant to a special needs school or classroom as explained in the following sections.

7.1.1 Literature Review

Chapter 1 outlined the difficulties and differences for autistic pupils with coexisting intellectual disabilities which may affect access to learning. The review of the literature explored how the interaction of autism and intellectual disabilities results in specific needs in these areas. A number of these difficulties were linked to diagnostic criteria such as communication difficulties, RRBs and sensory needs (APA, 2013; WHO, 2015). Further aspects such as the uneven cognitive profile, adaptive behaviour difficulties and behaviour that challenges may manifest in specific ways for this group or may be particularly prevalent with this population. The initial definition of this construct included broad areas or domains which may affect autistic pupils' opportunity to access and engage in learning.

As the above definition may include factors which are not specific to autism, such as socio-economic status, it was necessary for the definition to be narrowed. Therefore, barriers to learning were defined as affecting opportunity to access and engage in learning in a special needs classroom or school and able to be identified, observed and measured by special needs teachers. It was identified that these difficulties may be represented by broad domains or by

specific behaviours or skill gaps. It was also recognised that barriers to learning may be related to positive adult outcomes and quality of life.

The systematic review incorporated these areas under three categories, adaptive behaviour, behaviour that challenges and autism-related behaviour. The systematic review in Chapter 4 slightly extended the definition of adaptive behaviour by including participation assessments if they were appropriate for a school setting, could be used by teachers and if the focus was on skills or abilities relevant to participation as opposed to measuring levels of participation.

7.1.2 Focus Groups

The definition was further refined by the data derived from the focus group discussions. The data analysis of the focus group discussions on barriers to learning divided the areas identified as barriers into two categories. The first category confirmed the link with aspects of the diagnostic criteria and referred to areas such as RRBs and sensory needs. The second category related to skill gaps which may affect further learning for this group of pupils, such as communication impairments and difficulties with emotion and behaviour regulation. In this second category, teachers also specified that barriers to learning often included prerequisite 'learning' skills which pupils needed to show before curriculum or more complex learning could take place. The teachers provided a number of specific examples of these areas, explained the way they related to learning and identified that "*nuances and subtleties in behaviour"* can sometimes indicate the most important aspects of progress that pupils make.

The focus groups further helped to define the barriers to learning construct by outlining that the skills and behaviours relevant in terms of educational assessment are the ones that can be taught by teachers and learnt by the pupils. For example, although a lack of vocal communication may be considered a barrier to learning for a pre- or minimally verbal pupil, in the present context, a pupil could be taught to use an individualised method of AAC to communicate. The teaching and learning of the individualised communication methods would therefore be the focus in relation to the construct definition as this is the skill that can be taught, observed and measured by the teacher. As well as teaching specific skills, this may also include teaching coping mechanisms, supporting tolerance and teaching replacement skills or behaviours.

As mentioned in Chapter 1, teachers in special schools often attempt to use and devise appropriate adjustments to the school environment in order to reduce and overcome environmental barriers to learning. The focus group participants did note, however, that the overall, long-term aim in schools is usually to ensure that pupils can generalise these coping mechanisms or replacement skills to the 'real world'. It could be argued that this viewpoint is in contradiction to the social model of disability and posits the pupils as the 'problem' rather than the social environment which inhibits their access. However, in practice, it may not be as simple as requiring the current state of society and the social environment to adapt. To illustrate this, it may be more difficult to ensure the safety of an individual during emotional responses involving self-injurious behaviour in an outdoor public space than in a home or school setting, for example. Similarly, physical aggression towards other members of the community may limit opportunities to engage in environments where others are likely to be present such as play parks and leisure centres. Relying on environments which have been heavily modified, in a number of instances, may limit pupils' future opportunities as well as impact upon their individual and family quality of life. Quality of life outcomes would be improved, for example, if pupils could learn how to safely and appropriately access community spaces or tolerate weather-appropriate clothing. Therefore, although teachers are often limited to the school environment when teaching and supporting pupils to overcome barriers to learning, it is important that pupils are able to make progress with skills and behaviours which improve outcomes and quality of life outside of school as well as in the classroom. The promotion of independence and learning, as well as adapting teaching for pupils to learn skills which will extend their opportunities, is in line with aims of the social model of disability, such as holding positive expectations along with improving outcomes, opportunities and quality of life for these individuals.

The construct was therefore defined as the following. Barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities consist of autism-related characteristics or atypical skill development which affect the individual's opportunity to access and engage in learning in the school or classroom environment. Barriers to learning can be overcome by teaching new skills or refining existing abilities, teaching replacement skills or behaviours, providing exposure to new experiences, teaching coping mechanisms and using appropriate environmental adjustments. The barriers may be consistently present (or absent) or they may be context dependent. Barriers to learning are likely to include skill gaps and behaviours which may negatively impact upon future outcomes and quality of life.

7.2 Purpose

Chapter 2 outlined different types, purposes and methods of assessment. It also explored how assessment in special schools differs from that in mainstream schools and considered some overlap with psychological assessment. The ABLE-Autism was developed as an autism-specific outcome assessment which identifies barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities. It was developed for use by teachers in special schools and is intended to show change and progress in these barriers to learning over time.

The assessment has a number of aims and uses from a school and teacher perspective. Firstly, the assessment allows special needs teachers to identify pupils' barriers to learning or formally and systematically record existing barriers which have already been identified by the teacher. It may help teachers ascertain specific skills, areas or behaviours which may benefit from individualised and targeted support, teaching and intervention. It then enables teachers to measure and record change in these areas. This will support teachers to evidence when a pupil is making progress in important and relevant skills and behaviours, even if progress is not being made in curriculum learning or learning recorded through statutory assessment. Recording change also allows teachers to identify and evidence regression or unexpected difficulties a pupil might be having which may support a teacher to recognise additional or wider issues which may be affecting learning. The overriding nature of the ABLE-Autism is ipsative rather than norm-referenced or criterion-referenced. Although the assessment could conceivably be used to compare pupils, the assessment has not been normed at this stage and there are no specified criteria in which pupils have overcome or 'mastered' barriers to learning. The focus of the ABLE-Autism is on the individual pupil and, therefore, the way the assessment is used and the information provided from the assessment data will very much depend on the needs of each pupil and the reasons that the teacher is using the assessment. The ipsative nature of the ABLE-Autism very much reflects the focus group discussion which talked around individualisation, personalisation and holistic approaches towards assessment. It also takes account of the idea that the same progress in the same skill may, in reality, be very different achievements for different pupils. By recognising the importance of the starting point of individuals and not focusing on comparisons, the application of the ABLE-Autism can be centred on the individual. Purposes which are outside of the scope of the assessment include to identify underlying *causes* of the barriers, to assess the teaching, environment or instruction, or to suggest particular interventions or strategies to teach the

relevant skills. This will be discussed further when identifying research limitations in Chapter 9.

Consideration was given as to whether the assessment would be considered formative or summative in relation to the discussion in Chapter 2. Formative assessment is usually informal and the teachers in the focus group identified that good teaching practice with these pupils involves 'knowing the pupils well' and being very aware of their educational and learning needs. The ABLE-Autism was, therefore, considered too systematic and formal to be purely formative. The ABLE-Autism was developed to be used half-termly or less regularly over longer time periods if required, which indicated that it is of a summative nature. Reflecting the previous discussion of the blurred distinction between the two assessment types in special education, it was determined that the ABLE-Autism could be used by special needs teachers to more formally identify and record barriers to learning which they may have recognised through their formative assessment practice. Formative elements of the assessment tool allow teachers to use it at regular points in the school year, or however often they deem necessary, to enable a holistic overview of the pupil's needs and consider areas of intervention or next steps for learning. The assessment could then help teachers present progress summatively, particularly for pupils who may not be able to access more formal curriculum learning due to their barriers. The ABLE-Autism could also be used to evidence summative judgements and support summative assessment such as end-of-year reports and annual reviews. It could contribute to school level pupil data by providing additional evidence of progress outside of curriculum assessment. With the purpose of identifying difficulties, supporting planning and target setting, and recording and providing summative evidence of pupil progress, the ABLE-Autism reflects the description of formative assessment by Rouse and Agbenu (1998) discussed in Chapter 2. As the usefulness of the assessment is a key consideration in this research, the ABLE-Autism is not specifically precluded from use as either formative and summative assessment and teachers are able to make this decision on how formally or informally the assessment is used depending upon their assessment requirements and the needs of their pupils.

Finally, the assessment name was abbreviated to ABLE-Autism. This acronym was selected to positively reflect the research focus of supporting pupils on the autism spectrum to make progress and overcome barriers.

7.3 Description

7.3.1 Items and Subscales

The ABLE-Autism has 70 items which were endorsed by special needs teachers during the Delphi exercise. It is divided into five subscales: Learning Behaviour Barriers (14 items), Restricted and Repetitive Behaviour Barriers (12 items), Emotion and Behaviour Regulation Barriers (16 items), Sensory Barriers (10 items) and Functional Communication Barriers (18 items). These subscales developed iteratively through a number of stages described in previous chapters including a review of the literature, a systematic review, focus groups and the Delphi exercise.

The first list of assessment items for evaluation in the Delphi study was created as a result of the focus group discussions. Many of these items were developed from teaching priorities and barriers identified by the teacher participants and the item content frequently overlapped with difficulties and challenges discussed in Chapter 1. These items reflected a number of different situations which might arise. For example, one focus group participant identified a teaching priority as *"to be able to be in a group or… just be able to tolerate somebody near you"* and this was reflected in items such as *RRBB7a [Pupil] will accept a peer engaging in a preferred activity alongside them*. Many of the items were developed using wording that teachers used in the focus groups. For example teachers discussed *"changes in routine"* with and without warning which led to items *RRBB1a [Pupil] will accept changes to a normal routine with some warning* and *RRBB1b [Pupil] will accept changes to a normal routine without warning*. For other items, teachers mentioned general skill areas (e.g. *"enabling pupils to regulate their behaviour"* and *"recognising their own emotions… how they can manage those big feelings in an appropriate way"*) which then were broken down by the researcher and discussed with the supervisors who agreed on relevant skills applicable to the classroom, school or everyday life (e.g. items included in the EBRB subscale). As mentioned in Chapter 6, items were worded positively in keeping with the positive research focus of this thesis and in response to the focus group participants' expressing their desire for framing the pupils and their progress positively in terms of both assessment and teaching. It is worth noting that, prior to the Delphi study, a much larger proportion of the draft items were expected to be amended and excluded as a result of the responses in the Delphi exercise, as has been suggested in scale development literature (Boateng et al. 2018). Although the Delphi study included some different participants to the

focus group discussions, the large number of draft items accepted may reflect the fact that the items presented in the Delphi study were formulated based upon the teacher comments and wording in the focus group discussions.

A copy of the assessment instructions, subscales, items and response options are included in Table 6. Please note that, although the wording is accurate, the assessment format has been adapted for inclusion in the thesis and certain aspects have been amended for illustrative purposes.

Table 6. *The Assessment of Barriers to Learning in Education – Autism (ABLE-Autism)*[3]

The Assessment of Barriers to Learning in Education – Autism

(ABLE-Autism)

Instructions

Pupil Information

Complete the Pupil Information section before beginning the assessment. The pupil name will be pulled through to the assessment items. Please complete all questions. Questions require either a typed response or selection of an answer from the drop down menu. If the questions require selection from the drop down menu, this will appear by clicking the small, grey arrow which appears to the right when the answer box is selected.

Subscales

There are five subscales; Learning Behaviour Barriers, Restricted and Repetitive Behaviour Barriers, Emotion and Behaviour Regulation Barriers, Sensory Barriers and Functional Communication Barriers. You can view and complete each subscale using the tabs at the bottom of the page. Click on the three dots to the right or left of the tabs to view more tabs.

Statements of Skills/Behaviour

Consider the statement about your pupil, select the box which corresponds with how often you have seen that pupil display that skill or behaviour, click on the small, grey drop down box and select Y from the dropdown menu. That box will then be selected.
To clear a selection, select the box and press the delete key on your keyboard. Please ensure that you have provided an answer for all questions.

Optional Boxes

There is an optional grey box below each statement where you can specify 'what that looks like for this pupil'. You may use this box to individualise the assessment by describing the skill or behaviour more specifically in relation to the needs and abilities of that individual pupil.

[3] Attribution-NonCommercial-NoDerivatives 4.0 International

There is also an optional comments box for each statement where you can give additional detail, e.g. the level of support that pupil currently needs.

Guidance for responses:

Never - You have never witnessed or observed that pupil performing that skill or displaying that behaviour.
Occasionally - You have once or twice witnessed that pupil performing that skill or displaying that behaviour.
Regularly - That pupil performs that skill or displays that behaviour around 50% of the time.
Usually - The pupil performs that skill or displays that behaviour most of the time but not consistently every time.
Always - That pupil consistently performs that skill or displays that behaviour.

Pupil Information

Pupil's name:	*Type pupil's name*
Pupil's sex:	*Select from the following [male] or [female]*
Pupil's age:	*Type pupil's age*
Date assessment is completed:	*Type date [DD/MM/YY]*
Name of person completing assessment:	*[Type name of person completing assessment]*
Assessment complete by:	*Select from the following [Class teacher] or [Teaching assistant] or [Other]*
Please rate the difficulty that this pupil faces in accessing learning in the classroom or school:	*Select from the following [High – many barriers to learning] or [Medium – some barriers to learning] or [Low – few barriers to learning]*

Learning Behaviour Barriers

		Never	Occasionally	Regularly	Usually	Always	Comments (optional)
LBB1.	[Pupil] recognises their own belongings						

Describe here what that looks like for this pupil (optional)							
LBB2.	[Pupil] will show an understanding that something belongs to another person						
LBB3.	[Pupil] will show an awareness of the need to 'put things away' and that objects have a place where they belong						
LBB4.	[Pupil] will independently follow a familiar routine						
LBB5a.	[Pupil] will independently transition across a room following instructions or routine						
LBB5b.	[Pupil] will independently transition around the school following instructions or routine						
LBB6.	[Pupil] will show a clear preference for types of objects/activities						
LBB7.	[Pupil] is willing to try a new or unfamiliar activity						
LBB8a.	[Pupil] will appropriately engage for the duration of a short, adult led activity						
LBB8b.	[Pupil] will appropriately engage in a group activity with peers						
LBB9.	[Pupil] will complete a short, simple activity independently						
LBB10.	[Pupil] will independently complete a series of activities with visual support						

LBB11.	[Pupil] will accept an adult providing help						
LBB12.	[Pupil] will accept waiting for a short time						

Restricted and Repetitive Behaviour Barriers

		Never	Occasionally	Regularly	Usually	Always	Comments (optional)
RRBB 1a.	[Pupil] will accept changes to a normal routine with some warning						
Describe here what that looks like for this pupil (optional)							
RRBB1b.	[Pupil] will accept changes to a normal routine without warning						
RRBB2.	[Pupil] will accept waiting before accessing a preferred object/activity						
RRBB3.	[Pupil] will accept a brief verbal or physical interruption to an activity they are engaging in						
RRBB4a.	[Pupil] will accept a preferred object/activity ending after some warning						
RRBB4b.	[Pupil] will accept a preferred object/activity having a time limit						
RRBB5.	[Pupil] will accept the usual/preferred options not being available						
RRBB6a.	[Pupil] will accept an adult engaging in a preferred activity alongside them						

RRBB6b.	[Pupil] will accept an adult making changes to a preferred activity alongside them						
RRBB7a.	[Pupil] will accept a peer engaging in a preferred activity alongside them						
RRBB7b.	[Pupil] will accept a peer making changes to a preferred activity alongside them						
RRBB8.	[Pupil] will attempt to 'problem solve' if the usual way of achieving something isn't working						

Emotion and Behaviour Regulation Barriers

		Never	Occasionally	Regularly	Usually	Always	Comments (optional)
EBRB1.	[Pupil] will comply with simple adult instructions/direction to do something						
Describe here what that looks like for this pupil (optional)							
EBRB2a.	[Pupil] shows an understanding of the meaning of 'stop' or 'no'						
EBRB2b.	[Pupil] will respond appropriately to 'stop' or 'no'						
EBRB3.	[Pupil] is able to be present safely in public/community areas with adult support						
EBRB4.	[Pupil] will stay with an adult/the group when transitioning						

EBRB5a.	[Pupil] will show dislike or aversion in an appropriate way						
EBRB5b.	[Pupil] will show distress or sadness in an appropriate way						
EBRB5c.	[Pupil] will show that they want something to stop in an appropriate way						
EBRB6.	[Pupil] will appropriately attempt to remove themselves from upsetting or distressing situations						
EBRB7.	[Pupil] will accept adult intervention when distressed or emotions increasing						
EBRB8.	[Pupil] will accept an adult attempting to help when they are hurt or in pain						
EBRB9.	[Pupil] is able to stop or reduce inappropriate behaviour with support from adults						
EBRB 10a.	[Pupil] is able to independently stop or reduce behaviour which may harm or damage property						
EBRB 10b.	[Pupil] is able to independently stop or reduce behaviour which may hurt or harm themselves						
EBRB 10c.	[Pupil] is able to independently stop or reduce behaviour which may hurt or harm others						
EBRB11.	[Pupil] can 'move on' and return to an activity/setting/person after an emotional or distressing situation once they are calm						

Sensory Barriers

		Never	Occasionally	Regularly	Usually	Always	Comments (optional)
SB1.	[Pupil] will recognise when they need sensory input						
Describe here what that looks like for this pupil (optional)							
SB2.	[Pupil] will request sensory input when needed						
SB3.	[Pupil] will accept a delay in sensory input						
SB4a.	[Pupil] will appropriately indicate that they are hungry						
SB4b.	[Pupil] will appropriately indicate that they are thirsty						
SB4c.	[Pupil] will appropriately indicate that they are hurt/in pain						
SB4d.	[Pupil] will appropriately indicate that they need the toilet						
SB5.	[Pupil] will accept wearing clothes/shoes appropriate to or necessary for the weather/setting						
SB6.	[Pupil] will accept and try a new sensory experience with support and encouragement from an adult						

SB7.	[Pupil] will modify sensory-seeking behaviour to be appropriate to the context with direction						

Functional Communication Barriers

		Never	Occasionally	Regularly	Usually	Always	Comments (optional)
FCB1.	[Pupil] will show awareness of an adult who is seeking to gain their attention						
Describe here what that looks like for this pupil (optional)							
FCB2.	[Pupil] will attend to an adult talking directly to them						
FCB3.	[Pupil] will show awareness of the emotions of others						
FCB4.	[Pupil] will show an awareness of waiting for their turn or taking turns with others						
FCB5.	[Pupil] will seek an adult when something is wanted/needed						
FCB6a.	[Pupil] will appropriately initiate interaction with an adult when something is wanted/needed						
FCB6b.	[Pupil] will appropriately initiate interaction with an adult for pleasure or enjoyment						

FCB6c.	[Pupil] will appropriately initiate interaction with an adult to share information						
FCB7.	[Pupil] will show pleasure or enjoyment during a favourable interaction with an adult						
FCB8a.	[Pupil] will appropriately initiate interaction directly with a peer when something is wanted/needed						
FCB8b.	[Pupil] will appropriately initiate interaction directly with a peer for pleasure or enjoyment						
FCB8c.	[Pupil] will appropriately initiate interaction directly with a peer to share information						
FCB9.	[Pupil] will show pleasure or enjoyment during a favourable interaction with a peer						
FCB10.	[Pupil] will request help appropriately						
FCB11.	[Pupil] will request/acknowledge an object/activity/person that is not present						
FCB12.	[Pupil] will spontaneously make requests appropriate to the context						
FCB13.	[Pupil] will make appropriate attempts to be understood						

FCB14.	[Pupil] will respond 'yes' or 'no' appropriately when asked a question						

7.3.2 Method of Assessment

A number of assessment methods discussed in Chapter 2 were considered for the ABLE-Autism including rating scale, checklist and direct observation. A Likert-type rating scale was chosen for the following reasons. Firstly, in line with many assessments used in special schools, the assessment is a respondent report measure meaning it can be completed without the pupil being physically present. The assessment can be completed by a teacher who knows the pupil well and, as no training or qualification is needed to use the assessment, there is an option for teaching assistants to also complete the assessment if required. Secondly, a rating scale was considered quicker and less onerous to complete than direct observation or performance assessment. In comparison to a dichotomous checklist (i.e. which indicates whether a skill or behaviour is present or absent), a rating scale allows for further detail to be provided by the respondent. Each item in the ABLE-Autism is rated on a five-point Likert-type scale according to how often the teacher has observed that pupil performing that skill or displaying that behaviour (*never*, *occasionally*, *regularly*, *usually* or *always*). This was considered more straightforward than using terms which teachers discussed as potentially ambiguous in the focus groups (e.g. 'developing' or 'mastered'). The assessment instructions give teachers guidance on the meaning of the terms used in the assessment. For example, the guidance indicates that *usually* means that "*the pupil performs that skill or displays that behaviour most of the time but not consistently every time*".

For the purposes of the assessment evaluation outlined in Chapter 8, pupils were assessed using all five subscales. However, in ordinary use, the subscales were designed with the potential to be completed and used independently depending on the needs and barriers of a particular pupil.

Six of the items retained in round one of the Delphi exercise received one score of 4 for comprehensibility, meaning one teacher rated them not at all easy to understand (Appendix 5.3). Although these items were included as they exceeded the 80% threshold for both questions with a median of 1, it is important that all teachers who may use the assessment at any stage or level of their career understand the items. Therefore, they were

specifically marked for further consideration in the evaluation study and will be discussed in Chapter 9.

7.3.3 Features

The benefits of additional assessment information being provided was recognised by Cronbach (1970) who stated, "preferably, in individual testing, both scores and descriptive information are recorded" (p.33). When discussing assessment, the focus group participants talked about how assessments often "*don't tell the full story*" and that supplementary information is often important for assessment, yet there is often nowhere to include or record important information. In response to these findings, the ABLE-Autism has optional comments boxes for each item where teachers can note any additional information which may be helpful to them. Additional information may include the reasoning or justification for the response or, for example, the amount of support currently required which was mentioned by a number of additional comments in the Delphi exercise.

The ABLE-Autism also features a 'what that looks like for this child' box where teachers can personalise the item depending upon the targets or needs of the individual pupils. Personalisation can include minor amendments or specification to the items such as providing context or incorporating particular personal targets. For example, for *LBB1. [Pupil] recognises their own belongings*, teachers may specify a pupil's individual target by writing "Charlie recognises his coat when collecting it for playtime." This aspect of personalisation was created in response to focus group participants requiring assessments to recognise the "*unique and individual*" needs of each pupil. This feature may affect the reliability and validity of the assessment and this will be discussed further in Chapter 9. An additional minor feature relating to personalisation is that the assessment will pull the pupil's name from the pupil information page through into the list of items so that each item refers to them individually rather than just describing them as 'the pupil'.

To account for the individuality of pupils and to allow for the holistic nature of the assessment, examples were not provided and certain language was not defined. Many items contained the word 'appropriate' for example. The meaning of 'appropriate' was purposefully not defined as what is considered appropriate for one pupil may not be appropriate for another. Some examples of the variations of what may be considered appropriate in the context of the assessment items are outlined in Table 7.

Table 7. *Examples of Variation of the Term 'Appropriate' for Individual Pupils*

Item	Pupil 1	Pupil 2
LBB8a. [Pupil] will appropriately engage in a group activity with peers	The teacher of this pupil responds 'always' for this item as they will sit alongside peers during a shared parallel play activity at a table or in circle time.	The teacher of this pupil responds 'always' for this item as they will listen to the contribution of their peers and suggest ideas when playing with peers.
EBRB5a. [Pupil] will show dislike or aversion in an appropriate way	The teacher of this pupil responds 'always' for this item as they now hand an item they don't want to play with to an adult or move the item away rather than throwing it.	The teacher of this pupil responds 'always' for this item as they will verbally tell an adult when they don't like an activity.
SB4d. [Pupil] will appropriately indicate that they need the toilet	The teacher of this pupil responds 'always' for this item as they will hand a toilet symbol to an adult when they need the toilet.	The teacher of this pupil responds 'always' for this item as they will put up their hand and verbally ask a teacher if they can go to the toilet rather than just getting up and leaving the classroom.
RRBB4a. [Pupil] will accept a preferred object/activity ending after some warning	The teacher of this pupil responds 'always' for this item as they will move on to a different activity after being given a visual one-minute sand timer to show the activity is ending and a five second countdown at the point of transition.	The teacher of this pupil responds 'always' for this item as they will finish up their playground game with peers when told that playtime is ending in five minutes.

In these examples, the way the skill is displayed may differ between pupils, however the skill that is being assessed is consistent. It is also important to note that the assessment items do not require responses to the degree that they preclude environmental adaptations. For example, as indicated in the example of *SB4d* in Table 7, pre- or minimally verbal pupils will be able to indicate, respond or communicate for the purposes of the assessment with their own preferred and individual methods of AAC, such as symbols or signing, and still be displaying the relevant skill. RRBB4a in Table 7 also provides an example of environmental adaptations which a pupil may need visual timer to understand an upcoming transition and

make this transition predictable and easier for them. This is, again, in line with the social model of disability discussed in Chapter 1.

7.3.4 Scoring

When the full version of the assessment is created using software, it is anticipated that scores and summary sheet will be provided automatically. For the assessment evaluation, scoring was completed manually and returned to the teacher. Teachers or teaching assistants may also be able to score the assessment independently, however were not asked to do so for the evaluation study in Stage 3 of this research in order to reduce burden on participants.

For each item, teachers select how often the pupil displays the described skill or behaviour (*never* – 4, *occasionally* – 3, *regularly* – 2, *usually* – 1, *always* – 0). A score of 0 for an item indicates an absence of barriers to learning. The scores are summed for each subscale (subscale scores) and, if all subscales are completed, for the assessment as a whole (total score). The items which represent a pupil's primary and secondary barriers to learning (rated *never* and *occasionally* respectively) are highlighted on the score summary sheet. The higher the subscale or overall score, the more barriers to learning a pupil shows. The score summary sheet also displays a percentage for each subscale (the pupil score divided by the total possible score multiplied by 100) for a quick, accessible and straightforward comparison between subscales. Although the assessment was not developed for comparison between pupils, the percentages may also show differences between pupils should this be helpful in particular circumstances.

On subsequent completions of the assessment, the score summary sheet will show changes in scores from the previous completions. Changes in individual items and overall scores/percentages are highlighted. Positive progress in relation to barriers to learning is highlighted green and increased barriers are highlighted red. In response to the focus group discussions, the ABLE-Autism is not linear and changes in specific skills and behaviours related to both positive progress or potential regression can be accounted for. Although the assessment can record regression in skills, which was an assessment feature which teachers in the focus groups recognised as necessary, the ABLE-Autism can also show that a pupil may be making overall progress even if regression in skills or behaviour is seen on individual items or within a particular subscale.

The first and subsequent score summary sheet can be found in Appendix 6.

Chapter 8. Study 3 – Assessment Evaluation

8.1 Introduction

The findings of the systematic review in Chapter 4 showed that few assessments included in the review were developed with input from and in collaboration with teachers even though this is important to ensure content validity. In line with the findings and recommendations of the systematic review, the ABLE-Autism was developed specifically for the needs of autistic pupils with intellectual disabilities and input from special needs teachers was provided for both the construct definition and the item selection in Stage 2 of the research. The findings of the systematic review also showed that many assessments are not evaluated for robust measurement properties with appropriate populations, in relevant settings or with teacher respondents. Stage 3 of this research aimed to evaluate a number of relevant measurement properties of the ABLE-Autism. This chapter will outline the pilot testing and describe the results of the validity and reliability evaluation of the assessment. As discussed in Chapter 2, there are a number of measurement properties that may be relevant to evaluate in the context of educational assessment. The test-retest reliability and internal consistency reliability of the ABLE-Autism were considered as well as convergent validity with the TAPS (Dang et al., 2017). In line with recommendations from previous stages of the research which requires the assessment to be useful for the purposes it was developed, teacher feedback was also obtained on the usefulness of the new assessment for the specified purposes.

8.2 Methods

8.2.1 Participants

A total of 48 pupils were assessed using a convenience sample of 22 respondents from 13 schools in England and Wales. The teacher respondents were recruited in a number of different ways. Headteachers or assistant headteachers of over 75 special schools in England were contacted by email and asked whether any of their teachers wished to participate. Information about the research was also circulated to headteacher groups as well as posted on social media. Twenty-one teachers and one higher level teaching assistant (HLTA) with responsibility for classroom teaching and assessment participated (for consistency, the HLTA will be referred to as a teacher for the purposes of this stage of the research). Pupils' ages ranged from 4-15 years old (mean = 10.4 years, median = 11 years), all had a diagnosis of

autism as confirmed by their teacher and all attended a special school. Male pupils accounted for 77.1% of participants (n = 37). The number of pupils that each teacher assessed ranged from 1-7.

8.2.2 Ethical Compliance

Stage 3, the assessment evaluation stage of the research, required participation from teachers and teaching assistants who assessed their pupils using the newly developed assessment. Teachers agreed to participate with permission from headteachers or another member of the school leadership team. As active participants, teachers were provided with an electronic copy of the teacher information sheet and consent form and were required to print, sign, scan and return the form to the researcher before the research could begin. In some schools, despite instructions to the contrary, the assessment was shared by one teacher to others before the consent form was signed. In these cases, consent was implied by the completion of the assessment and the teacher or teaching assistant was sent reminders until the consent form was completed and returned retrospectively. All participating teachers provided written consent. As with the focus groups and Delphi exercise, the teacher information sheet explained the research background, detailed account of the assessment procedures, the voluntary nature of participation, the right to withdraw, confidentiality and information about data storage.

It was also necessary to consider issues of consent relating to the pupils who were to be assessed. Although opt-out consent is not usually considered appropriate for research with children, for assessment of the pupils in this instance, opt-out consent by parents was initially considered appropriate for a number of reasons. Firstly, it is not usually necessary for parents to consent to assessment systems deemed appropriate by the school. The parents and children were not asked to actively participate and, instead, active participation in the research was by teacher participants. Secondly, parents who are most responsive to school communication could also be particularly engaged in their child's education. This may have resulted in the inclusion of pupils who have the least barriers to learning due to the consistent use of educational strategies across school and home settings, resulting in skewed data. Thirdly, it was not necessary to know the identity of the children who are subject to the assessment tool by their teachers. Data were required on pupils' gender, age and/or school year/key stage but their exact date of birth, name or any other personally identifiable information was not required. Data passed to the researcher were anonymised by the teacher and contained no

information from which the pupil can be identified. Finally, some communication channels from schools to parents are indirect, e.g. providing a consent form and information sheet into a pupil's home-school communication book. The pupils that teachers identified as potentially benefitting most from a new assessment in these areas may not have been included if parents did not respond to these indirect communication methods even if they were happy and willing for their child to participate. Parents of pupils that the teachers chose to assess were provided with parent information sheets containing the same information as the teacher information sheet in simple and jargon-free language. Opt-out consent forms along with the information sheets were sent to parents via the schools' communication methods. Parents were able to opt their child out of this research by a variety of means including returning opt out forms to the school, speaking directly to the school office or class staff or contacting the researcher on the contact information provided. An appropriate time frame of at least two weeks was given and made explicit on the information sheet in order to give parents time to opt-out before the research began. Contact details of the research team were provided in case teachers or parents had further questions about the study. It was made explicit on the teacher and parent information sheets that teachers and parents were able to withdraw from participation at any time. One headteacher whose teachers agreed to participate specified that they required opt-in consent for parents and an ethical amendment was approved on 20/11/2019 allowing for parent opt-in consent forms and information sheets. Informed, written consent was received from the parents of 43 participating pupils. The remaining five pupils received opt-out consent forms and parents of all five of these pupils indicated to teachers through other methods (e.g. in the home-school communication book, written, verbally) that they were happy for their children to be involved. Pupil information was anonymised and pseudonymised by the class teachers before the assessment results were sent to the researcher so that no personally identifiable information was shared.

8.2.3 Procedures

The ABLE-Autism was completed by the pupils' class teacher during the autumn and/or spring term of the 2019/20 academic year. Each teacher had known the pupil since at least September and had been teaching that pupil for a minimum of ten weeks at the time of the first assessment. The teachers were provided with information about the assessment and instructions for the first stage of the research. They were asked to conduct an initial completion (Time 1) and a second blind completion (Time 2) approximately two weeks later. Teachers were also invited to complete the TAPS at Time 1. The initial study information

required the teachers to assess the pupils a third time approximately six to eight weeks after the initial completion in order to evaluate the responsiveness of the assessment. However, this final stage of the assessment evaluation with a focus on responsiveness could not be conducted as the timing of all but two teachers' final completions fell during the COVID-19 pandemic which resulted in school closures or limited attendance at schools in the United Kingdom. Teachers were also asked to complete a feedback questionnaire in April, 2020, on completion of the data collection.

8.2.4 Teacher Autism Progress Scale

The TAPS was identified in the systematic review in Chapter 4. It was one of the few assessments identified as potentially being appropriate for the research purposes as it is an autism-specific, teacher assessment developed to measure progress for pupils on the autism spectrum in behaviour, social abilities and functional skills (Dang et al., 2017). It was also developed in collaboration with teachers and researchers and is designed to be completed weekly to show small changes in pupil progress. The TAPS differs to the ABLE-Autism in that it is shorter, scored in the opposite direction, intended to be conducted more frequently and does not specify the level of detail within items as the ABLE-Autism. The TAPS has 16 items and teachers are asked to rate how often the pupil has engaged in the described activities or behaviours in the past week on a six- or seven- point scale (e.g. *never, sometimes, occasionally, often, usually, almost always, always*). One item requires teachers to select in which ways, if at all, the pupil has shown aggressive behaviours. Higher total TAPS scores indicate better performance (i.e. fewer difficulties). Previous research has shown a statistically significant negative correlation between TAPS scores and scores on both the Social Responsiveness Scale (SRS; $r = -0.70$; $p < 0.001$) and the Aberrant Behaviour Checklist (ABC; $r = -0.50$; $p = 0.008$) (Dang et al., 2017). Statistically significant improvements in TAPS mean scores were found over time, while improvements in the SRS and ABC mean scores were non-significant (Dang et al., 2017). Permission to use the TAPS in the current study was sought and granted. Although further validity and reliability evaluations of the TAPS have not been conducted, the TAPS was chosen as an assessment for comparison with the ABLE-Autism for three reasons. Firstly, the TAPS was similar to the ABLE-Autism in the areas assessed, was autism-specific, could be completed by teachers and was developed with input from teachers to be used in schools to show progress. Secondly, the responsiveness of the TAPS to show progress of autistic pupils in schools had been previously evaluated. Although the responsiveness element of the ABLE-Autism was not

able to be evaluated because of school closures due to the COVID-19 pandemic, initially this evaluation was considered an important part of this study. Finally, the TAPS was selected as it is a short assessment and this was considered necessary to avoid adding to the workload of teacher participants and to reduce the likelihood of participant drop-out.

8.2.5 Data Analysis

Although there is disagreement among psychometricians, psychologists, and sociologists on how to treat Likert-type scale data (Sullivan and Artino, 2013), Likert-type scales are widely treated as interval level measurement for evaluation purposes, particularly pilot testing (Furr, 2011). Limiting data analysis of scales to nonparametric approaches may be overly restrictive when studies have suggested that parametric analysis of scale data can be meaningful and appropriate (Harpe, 2015). Likert-type scales are often created to measure an underlying continuous variable (Allen and Seaman, 2007) and research has provided evidence that parametric tests can be robust for the analysis of summed Likert scale scores, even with non-normal distribution and small sample sizes (Queen et al., 2002; Murray, 2013; Wadgave and Khairnar, 2016). Parametric methods of analysis were therefore considered appropriate for this research.

8.2.5.1 Normality

Adjusted composite and subscale scores from Time 1 were assessed for normality through a combination of visual inspection of histograms and Q-Q plots, assessment of the skewness and kurtosis values and evaluation using the Shapiro-Wilk test. Adjusted composite and subscale scores all appeared approximately normally distributed apart from the Sensory Barriers subscale. The Sensory Barriers subscale score had skewness, kurtosis and Z values all within the range for normality for small samples, however the Shapiro-Wilk test indicated a small departure from normality with slight negative skewness (p = .032). In practice, data often have slight departures from normality and, in this instance, the skewness and kurtosis Z values suggested deviation is not marked and the use of parametric tests was considered meaningful and appropriate. Skewness, Kurtosis, Z values and the Shapiro-Wilk Test are shown in Table 8.

Table 8. *Skewness, Kurtosis, Z Values and Shapiro-Wilk Test for Adjusted Subscale Scores and Total Scores*

Adjusted Scale/Subscale	Skewness	Skewness Z Value (Skewness/Std Error)	Kurtosis	Kurtosis Z Value (Kurtosis/Std Error)	Shapiro-Wilk Test Sig.
Adjusted LBB	-0.012	-0.035	-0.801	-1.188	.540
Adjusted RRBB	0.155	0.452	-0.678	-1.006	.676
Adjusted EBRB	-0.075	-0.219	-0.917	-1.361	.188
Adjusted SB	-0.549	-1.601	-0.574	-0.852	.032
Adjusted FCB	0.001	0.003	-1.112	-1.65	.106
Adjusted Total	0.003	0.009	-0.967	-1.435	.277

As explained in Chapter 3, multilevel modelling was used to account for the nesting of data with pupils as the unit of analysis (level one) and teachers as the nesting variable (level two). Statistical analyses were conducted using IBM SPSS Version 26.

8.2.5.2 Missing Data

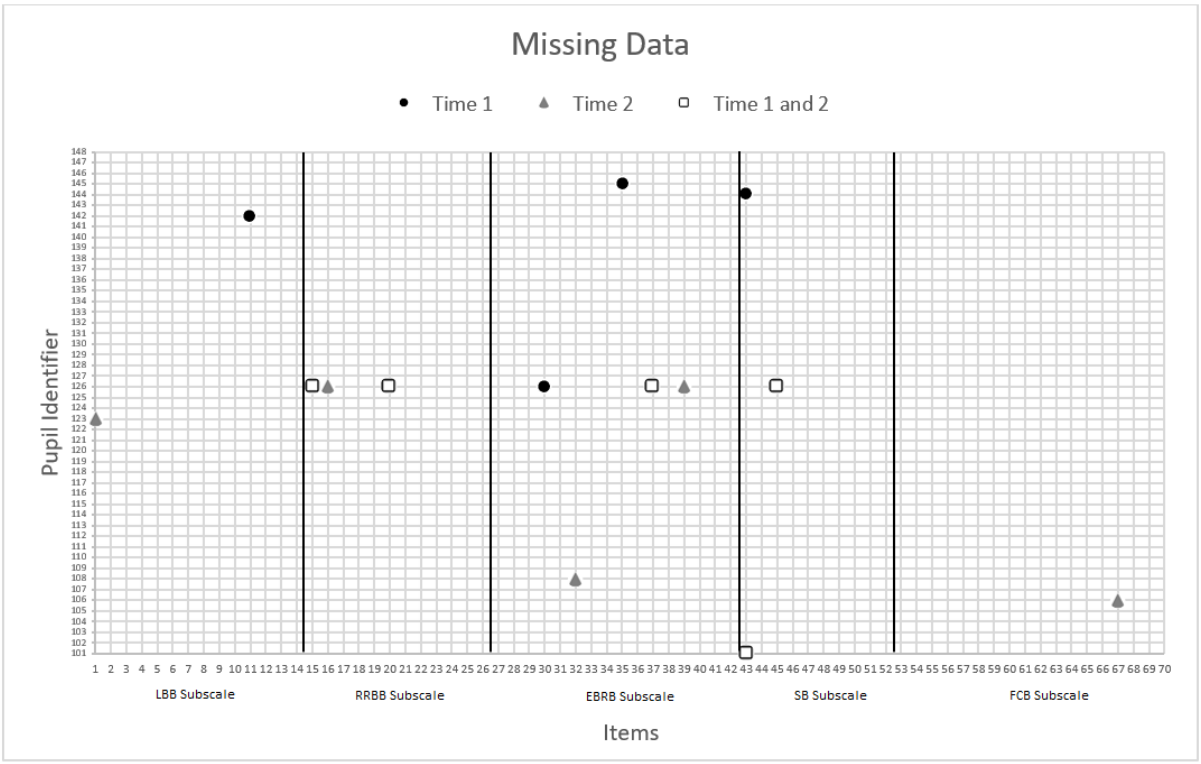
In total, nine item responses were missing at Time 1. One item had two missing responses at Time 1 (*SB1*) with the rest having just one missing response (*LBB9, RRBB1a, RRBB4b, EBRB3, EBRB6, EBRB8, SB3*). The assessment of one pupil had five missing responses with four pupils having just one missing response each.

Ten responses were missing at Time 2 (*LBB1, RRBB1a, RRBB1b, RRBB4b, EBRB8, EBRB10a, EBRB5a, SB1, SB3, FCB11*). No items at Time 2 had more than one missing response. The assessment of one pupil had six missing responses with four pupils having just one missing response each. One teacher respondent had one missing response for two pupils who were assessed. The number of missing items was < 1% of all items completed at both Time 1 and Time 2 and the average missing items per completion was also < 1%. Five items had missing responses at both Time 1 and Time 2. The data are described in Table 9 and the pattern of missing items is shown in Figure 3.

Table 9. *Numbers of Missing Items*

1st Completion	2nd Completion
Total items n = 3,360	Total items n = 2,450
Missing item responses n = 9	Missing item responses n = 10
Items with a missing response n = 8	Items with a missing response n = 10
Pupils with at least one missing item n = 5	Pupils with at least one missing item n = 5
Respondents with at least one missing item n = 5	Respondents with at least one missing item n = 4
Average missing items per completion n = 0.19	Average missing items per completion n = 0.29

Figure 3. *Scatterplot of Missing Data*.



In the absence of a not applicable option, missing item responses were noted on score summary sheets which were returned to teachers and these items were allocated a score of 0 (equivalent to a response of *always*). A score of 0 was allocated because, for the purposes of this assessment, it was not considered appropriate to infer barriers which were not identified by the respondents. If the item was not completed because the teacher felt the item was not

applicable to the student, then 0 is the most relevant score suggesting that skill or behaviour is not a barrier for that pupil. Similarly, if an item was accidentally missed, it was not appropriate to make an assumption that the item constitutes a barrier for that pupil.

Missing data in the context of the data analysis were dealt with by pairwise exclusion for internal consistency analysis. For convergent validity and test-retest reliability analysis, total scores and subscale scores were adjusted to account for missing item responses.

8.2.5.3 Internal Consistency

Internal consistency of each of the subscales and of the whole assessment as a unidimensional scale was determined by calculating Cronbach's α coefficient. There are some limitations of Cronbach's α when calculating internal consistency and alternatives were considered (Trizano-Hermosilla and Alvarado, 2016). However, due to the small sample size and academic disagreement on the best placed alternatives, Cronbach's α was used to calculate internal consistency in this instance with acknowledgement of its potential limitations (Sijtsma, 2009). In the absence of a sample large enough to conduct a factor analysis, it is necessary to note that internal consistency evaluations cannot demonstrate that the items all measure the same latent construct (Leppink and Perez-Fuster, 2017). Calculating Cronbach's α relies on the assumption that the scale or subscale is unidimensional and, therefore, α can be used only to supplement information about the factor structure of a scale, rather than provide evidence for it. As factor analysis was unable to be conducted in this study due to sample size, α was calculated on both subscales and the scale as a whole as a precursor for future factor analysis. Internal consistency was calculated on the first completion for each pupil which included data from a total of 48 assessments conducted by 22 teachers. Cronbach's α was calculated accounting for the use of dependent data using a three-level model (Level 1 – item level, Level 2 – pupil level, Level 3 – teacher level) outlined in Nezlek (2017).

$$\text{Item level reliability: } \alpha = \frac{\sigma^2{}_{\text{pupil level}}}{\sigma^2{}_{\text{pupil level}} + \frac{\sigma^2{}_{\text{item level}}}{\text{p}}}$$

The COSMIN Manual suggests good internal consistency is indicated by an α value ≥ .70 (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018).

8.2.5.4 Test-Retest Reliability

Test-retest reliability is considered important in the evaluation of new measures as it is the only way to show how similar the results are when an assessment is repeated with the same participants (Leppink and Perez-Fuster, 2017).

Eleven teachers conducted a blind completion of the assessment a second time for 35 pupils approximately two weeks (range = 6-30 days, median = 14 days, mean = 14.37 days, SD = 5.07 days) after the initial completion. An appropriate time interval between the test administrations for test-retest reliability analysis will depend on the specific assessment, purpose and context (Leppink and Perez-Fuster, 2017). In the current study, because the assessment was intended to reflect progress over a half term (i.e. potentially over a six to eight week period), a short period over which to conduct test-retest reliability was considered appropriate. The number of items and the fact that over half of the participating teachers were assessing a number of pupils each meant that recall effects were minimised.

Intraclass correlation coefficients (r_i) were considered to report the test-retest reliability of the assessment. However, traditional methods of calculating r_i do not take nesting of data into account. An approach to determine effect sizes in multilevel models was outlined by Lorah (2018) and was therefore chosen to account for the nested data in the current study. The effect sizes related to variance explained for the multilevel random intercepts models were reported using values for R^2, f^2 and r_{es}. Firstly, R^2 was calculated to determine the variance explained at the teacher level, followed by f^2 which represents the variance explained at the teacher level relative to other levels. This was transformed to a correlation coefficient, r_{es}, for easier interpretation. The values for f^2 and r_{es} were calculated and reported for both adjusted subscale scores and the adjusted total scores. Comparisons with random slopes models were considered, however the effect sizes for the random slopes models could not be calculated due to model non-convergence and therefore the random intercepts models were used in the present study. The relevant formulas outlined by Lorah (2018) are shown below.

$$R^2 = 1 - \frac{\sigma_F^2 + \tau_F^2}{\sigma_E^2 + \tau_E^2}$$

$$f^2 = \frac{R^2}{1 - R^2}$$

8.2.5.5 Teacher - Teaching Assistant Reliability

When considering how to evaluate inter-rater reliability, it was clear that in special schools there was unlikely to be another qualified teacher who knew the pupils to the same degree as the class teacher. As recognised in the focus groups, the key to effective teaching and learning for teachers in special schools is knowing their pupils well and, unless two teachers job share a class teacher role, there is unlikely to be an equivalent teacher who understand the needs of the pupil in the same way as the class teacher. Classes in special schools often have a number of teaching, learning or support assistants (TAs) who work in the class every day and also have a good understanding of the pupils' needs. It was recognised that, although the TAs are unlikely to have the same knowledge, experience and training as qualified class teachers, special schools often require TAs to contribute to aspects of teaching and assessment. Some TAs may therefore have experience or interest in assessing pupils and, for that reason, participating teachers were asked if any class TAs may wish to participate by completing the assessment separately to the teacher. Unfortunately, only four TAs were interested in participating and, subsequently, there were not enough data for analysis.

8.2.5.6 Convergent Validity

Class teachers of 41 pupils also completed the TAPS within approximately two weeks of their first completion of the ABLE-Autism (range = 0-21 days, median = 4 days, mean = 6 days) in order to evaluate convergent validity. To account for teacher level variance, the same method as for test-retest reliability was used to evaluate the convergent validity between the total score and TAPS total score (Lorah, 2018). The values for f^2 and r_{es} were calculated and reported for the adjusted total score correlated with the TAPS total score. It was hypothesised that the correlation between the adjusted total score of the ABLE-Autism compared with the TAPS was likely to be medium to high.

8.2.5.7 Teacher Feedback

In order to triangulate data from Stage 2 of the research and to ensure the assessment was useful to teachers, participating teachers were asked to complete a short feedback

questionnaire at the conclusion of the data collection. Teachers could either complete the questionnaire anonymously online via Qualtrics or return the questionnaire by email (Appendix 6.5). The questionnaire contained four questions asking teachers about the comprehensibility, relevance, comprehensiveness and usefulness of the assessment as suggested by the COSMIN manual and in line with the Delphi exercise. Teachers answered the four questions on a 10-point scale ranging from 1 (e.g. not at all useful, not at all relevant) to 10 (e.g. extremely useful, extremely relevant). A comments box was provided where teachers could choose to add additional comments to any of the questions to give an insight into the responses and to ensure that reasons for scores could be determined. A fifth question asked teachers they had any further information they would like to share, with space to comment on aspects of the assessment that they liked or found useful as well as aspects of the assessment they think could be changed or improved. A final box provided space for any further comments. A total of 16 teachers responded to the feedback questionnaire giving a response rate of 73%.

8.2.5.8 COVID-19

COVID-19 was first identified in the United Kingdom on 31/01/20 which was near the beginning of the Spring term and resulted in widespread school closures from 20/03/20. A number of participants were due to begin their second completion around the time schools were closed with some participants nearing the timescale for their third completion. An important part of the original study design involved an evaluation of the responsiveness of the ABLE-Autism to provide evidence that the assessment captured change over time. Evaluating the responsiveness of the assessment was considered important in the current study as "an instrument cannot be used to detect a change when its measurement error is larger than the change we want to be able to detect" (Berchtold, 2016, p.2). It was important to ensure that any change that the assessment may show over time was measuring real pupil progress or regression rather than as a result of measurement error. The third stage of the research was not able to be carried out as, even when pupils returned to school, educational disruption was ongoing throughout 2020. As a result of the school closures, the number of second completion responses were affected as well as the possibility for many teaching assistants to complete the assessment. Some teachers who had begun the second completion or who had pupils still attending school completed the assessment the second time shortly after the school closures were first announced.

8.3 Results

8.3.1 Descriptive Statistics

Descriptive statistics of the Time 1 adjusted scores including mean, median, standard deviation, minimum and maximum are shown in Table 10.

Table 10. *Descriptive Statistics of Adjusted Total and Subscale Scores at Time 1*

Adjusted Scale/subscale	Mean	Median	SD	Minimum	Maximum
Adjusted LBB	1.98	1.97	0.91	.14	3.79
Adjusted RRBB	1.93	1.88	0.80	.33	3.58
Adjusted EBRB	2.02	1.84	0.89	.25	3.63
Adjusted SB	2.24	2.50	1.01	.00	3.80
Adjusted FCB	2.25	2.31	0.83	.83	3.78
Adjusted Total Score	2.08	2.06	0.75	.71	3.60

8.3.2 Internal Consistency

Cronbach's α was calculated for each subscale and the total scale using the three-level model described above. Values of α for all subscales as well as the total scale were above .70 which is considered to be the threshold for good internal consistency (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). Values of α are reported in Table 11.

Table 11. *Values of α for Subscales and Total Scale Using a Three-Level Model*

Scale/subscale	α
LBB	.89
RRBB	.86
EBRB	.87
SB	.80
FCB	.88
Total Scale	.95

8.3.3 Test-Retest Reliability

Based on the formulas described above, f^2 and r_{es} were calculated for each adjusted subscale score and the adjusted total score. Values for f^2 and r_{es} could not be calculated for the LBB subscale due to model nonconvergence. The single measures, two-way mixed r_i with absolute agreement definition was reported instead along with the 95% confidence interval as suggested by Koo and Li (2016). The r_i for this subscale, however, does not take account of the nested data. Values are reported in Table 12.

Table 12. *f^2 and r_{es} for Adjusted Total Score and Subscale Scores*

Scale/subscale (adjusted scores)	f^2	r_{es}
*LBB***	Not calculated due to model nonconvergence r_i = .97 (95% CI 0.94 - 0.99, p < .001)	
RRBB	7.92	0.94
EBRB	5.92	0.92
SB	5.94	0.93
FCB	14.75	0.97
Total Score	13.10	0.96

Note. ** r_i two-way mixed, absolute agreement, single measures reported (does not take account of nested data)

All f^2 values can be interpreted as showing a large effect ($f^2 \geq .35$; Lorah, 2018) and the r_{es} correlation coefficient accounting for effect size is also high for all subscales and the total score. The r_i for the LBB subscale is excellent ($r_i > 0.9$ indicate excellent reliability; Koo and Li, 2016). Test-retest reliability was therefore shown to be excellent for all subscales and the total score.

8.3.4 Convergent Validity

The formulas described above were also used to calculate f^2 and r_{es} to determine the convergent validity with the TAPS based on data from 41 pupils. The value of f^2 was calculated as 1.74 and r_{es} = 0.80. The results suggested that scores on the two assessments had a strong correlation, with high ABLE-Autism scores (indicating greater difficulties) correlating with low TAPS scores (indicating greater difficulties). This suggested that the ABLE-Autism and the TAPS appear to be assessing similar constructs, providing evidence for convergent validity.

8.3.5 Teacher Feedback

Teacher feedback was considered to be a key part of the research process. As acknowledged throughout this thesis, an assessment may have sound measurement properties but it is essential that it is also considered useful by the teachers who will use it. Teachers were therefore provided with a feedback questionnaire containing the questions outlined in Table 13. Teachers were also given the opportunity to supplement their answers with further comments if they wished. Questions were similar to those which were asked in the Delphi exercise and covered key areas of content validity including relevance, comprehensiveness and comprehensibility (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018).

As can be seen in Table 13, on a 10-point scale where 1 was low and 10 was high, mean scores for all four questions were over 8 with median scores of 8 or 9. This initial feedback was extremely positive and these encouraging results may reflect the fact that special needs teachers were consulted at every stage of the assessment development process. This will be discussed further in Chapter 9.

Table 13. *Descriptive Statistics of Teacher Feedback Scores.*

Question (1 Not at all – 10 Extremely)	**Mean**	**95% Confidence Interval**	**Median**	**Standard Deviation**
Q1. How easy were the descriptions of the skills/behaviours to understand?	8.56	7.69-9.43	9	1.63
Q2. How relevant were the skills/behaviours to your pupils who show barriers to learning?	8.5	7.72-9.28	8	1.46

Q3. Were the assessment and score summary sheets useful in highlighting progress and/or areas of concern?	8.13	6.79-9.46	9	2.5
Q4. Would you find the assessment useful to assess pupils who do not appear to be making progress in the curriculum (either using the whole assessment or any of the subscales alone)?	8.06	6.78-9.35	8	2.41

Optional comments were provided by 10 teachers and these ranged from comments which confirmed their responses (e.g. "*clear to understand what all descriptors meant*") to comments about the physical use of the assessment (e.g. "*the drop-down button for highlighting a cell is far too time consuming*") and suggestions of additions (e.g. "*expanding / adding a little more about emotional regulation and sensory needs*") or changes (e.g. "*a N/A box might have helped*"). Some participants also provided general comments on usefulness (e.g. "*both the assessment and the score summary sheet are very clear/visual and will be extremely useful to use*").

8.4 Discussion

8.4.1 Reliability and Validity

The results of the data analysis provided preliminary support for the validity and reliability of the ABLE-Autism. The findings indicated that the individual subscales and the entire assessment had a high degree of internal consistency. Cronbach's α is known to be higher for assessments with large number of items and an extremely high α value may indicate that some items are redundant (Tavakol and Dennick, 2011). However, as the importance of items was determined by teachers in the Delphi exercise, it is necessary to consider the relevance and usefulness of the information provided by individual items along with the statistical analyses of the assessment's measurement properties. Therefore, potentially useful items were not considered for removal based on α values. As mentioned above, it is important to note that a high α value does not suggest unidimensionality and factor analyses will be required in order to determine the underlying factor structure. This preliminary internal consistency analysis will lend support for the internal structure of the assessment determined by future factor analysis.

Test-retest reliability was shown to be high, with r_{es} values and the r_i for the LBB subscale all falling within the excellent range. The data provide evidence that the test items are specific enough to yield the same answer at different administrations. The r_{es} values for test-retest reliability may be high because class sizes at special needs schools are usually small and, therefore, the teachers often know their pupils particularly well. This may mean that teachers are aware of individual pupils' skills, abilities and behaviours at a given time and are consistent in their assessment of the pupils. The high r_{es} values may also be suggestive of a short time interval resulting in recall effects. However, as there are 70 items in the assessment, recall effects are likely to be minimal and the two-week test-retest interval was sufficiently short to ensure that scores were unlikely to be impacted by developmental change.

Although a measure cannot be valid without being reliable, it can be found to be reliable without being valid. In addition to the face and content validity considered during the development of the assessment through the Delphi exercise, validity was further evaluated by determining the convergent validity with the TAPS. Similar to the ABLE-Autism, the TAPS assessed areas which may impact upon the ability of classroom and school engagement and learning for pupils on the autism spectrum including functional communication, emotion and behaviour regulation, attention, focus and levels of support. The TAPS, however, is a much shorter assessment covering broader areas whereas the ABLE-Autism focuses on smaller and more specific aspects of these skills and behaviour. The TAPS also considered how often the teacher has seen the pupil display the behaviour in the last seven days whereas the ABLE-Autism asks teachers to draw upon wider knowledge of that pupil. A moderate to strong correlation between the ABLE-Autism total score and the total scores on the TAPS was therefore hypothesised. The high r_{es} value provided evidence for the validity of the ABLE-Autism and suggested that it assesses areas which teachers identify as potential difficulties for pupils on the autism spectrum in a school environment.

Further evidence for validity could be provided through exploratory or confirmatory factor analysis, however the sample size in this initial evaluation research precluded factor analytic methods (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). It is recommended that factor analysis be carried out in future studies in order to determine the uni- or multidimensionality of the scale and subscales.

8.4.2 Use of the Assessment and Teacher Feedback

Teacher feedback was extremely positive and suggested that the assessment was used as intended by the participating teachers. With mean and median scores above 8 for all four questions, it is clear that the majority of the teachers who responded found the assessment easy to understand, relevant to pupils on the autism spectrum, useful to show pupil progress and comprehensive in the skills and behaviours it assessed.

When considering teacher feedback, it is necessary to acknowledge that, although teachers may have agreed to participate in the trial of the assessment due to the needs of particular pupils they worked with, teachers were still using the assessment artificially for purposes of this research. Teachers were limited, for example, to using the assessment with pupils whose parents consented for them participate and were also asked to complete all subscales regardless of whether they considered their pupils to show barriers to learning in the different areas. The teachers also completed the assessment on top of their classroom duties and usual pupil assessments. Teachers choosing to use the assessment outside of this research are likely to select the assessment due to the needs of their pupils and the potential lack of appropriateness of other curriculum assessments for individual pupils. They also may choose to use individual subscales which are particularly relevant to the learning barriers of their pupils. Teachers would not ordinarily need to complete the whole assessment under normal circumstances if they did not consider it appropriate to do so.

When considering the optional teacher comments, 28 individual comments remarked positively on the usefulness of the assessment with four individual comments stating that the assessment was easy to understand and four further comments expressing its relevance to the pupils. A number of teachers suggested additional areas which could be included in the assessment (e.g. self-help skills) and one teacher suggested that the assessment may be useful for parents to complete to provide a new teacher with information about the pupil. Five further comments by teachers related to the practical features of the assessment such as clicking buttons to make selections. Two teachers suggested that an N/A box would be useful. As the assessment used by teachers in this study was a prototype, comments on features and design of the assessment will be taken into consideration when the final format of the assessment is created. Only two teachers made comments that the assessment did not add to their current assessment practices and these comments may, as mentioned above, reflect the fact that there was a degree of artificiality in using the assessment for the research

purposes. All other teachers who left optional comments suggested that the assessment would be useful, either wholly or partly, with some or all of their pupils.

8.5 Limitations

The preliminary pilot testing of the ABLE-Autism was positive and provided initial evidence of the validity and reliability of the new measure. There are, however, some limitations that need to be considered and addressed in future research.

Firstly, as with most pilot tests of new measures, a larger sample is always preferable. Johanson and Brooks (2010) explored sample sizes for preliminary pilot studies on initial scale development and suggested a minimum sample size of 30 for these purposes and the sample used in this study is therefore considered adequate for this initial evaluation. The sample size did, however, preclude further validation such as principal component analysis, factor analysis and standardisation.

Secondly, the autism diagnosis of participating pupils was not checked or confirmed (i.e. through administration of diagnostic assessment). Teachers and parents were told that only pupils with a diagnosis of autism could participate and, although schools and teachers were relied upon to only select pupils who had an autism diagnosis, it may be that pupils with autistic characteristics who were members of autism-specific schools or classes but without a formal diagnosis were included in the study.

Thirdly, although the TAPS which was used for convergent validity evaluations had been found to have adequate quality evidence of sufficient responsiveness, aspects of reliability and validity had not been evaluated and evaluations had been conducted in the United States rather than the United Kingdom. Therefore, further convergent validity evaluations with relevant measures are necessary in order to provide further evidence for the validity of the ABLE-Autism.

Additionally, teachers were provided with a blank assessment for the second completion and instructed to complete it blind without reference to their previous responses. As the assessments were completed remotely and returned to the researcher electronically, teachers did have access to their previous responses when completing the assessment for the second time and it is a possibility that some teachers referred to previous responses even though they were instructed not to.

Finally, teachers completed the assessment for this preliminary study under different circumstances than if pupils were assessed based on need. Teachers and individual schools have a degree of autonomy as to the assessments they use with different pupils but, in this study, teachers were limited to assessing pupils whose parents had consented rather than pupils who they had specifically chosen to assess. Teachers may therefore have assessed pupils who have minimal barriers to their learning, who were accessing curriculum content and appropriately being assessed through other school assessments.

It is necessary for these limitations to be considered in subsequent validation of the assessment to ensure that the assessment is further evaluated in appropriate contexts.

8.6 Conclusions

The preliminary evaluations of the new ABLE-Autism are promising. The assessment shows convergent validity with the TAPS, excellent internal consistency and good to excellent test-retest reliability. Of equal importance is that the ABLE-Autism received extremely positive feedback from the teachers who used it. An idea central to this thesis is that an assessment needs to be practically useful for the purposes for which it was developed. The positive responses received from teachers in the feedback questionnaire suggested that, along with being valid and reliable, the ABLE-Autism is both useful to teachers and relevant to pupils on the autism spectrum with intellectual disabilities. The implications of the reliability and validity evaluation, along with the teacher feedback and recommendations for future research, will be further explored in Chapter 9.

Chapter 9. Discussion

9.1 The Research Question

This thesis aimed to address the following research question:

> How can the barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities be effectively identified and monitored through robust assessment by teachers in special schools?

This research question was broken down into a number of more specific research questions which were addressed by each of the four studies. The findings of each study led to the subsequent stage of the research and resulted in the development of the ABLE-Autism; an assessment for teachers in special schools to identify, monitor and show progress in barriers to learning for autistic pupils with intellectual disabilities. In this chapter, the results, outcomes and implications of the individual studies will be discussed in relation to each other, the development of the ABLE-Autism and the literature outlined in Chapters 1 and 2. The theoretical and practical implications of this research will be considered as well as a further discussion of the limitations of the studies and recommendations for future research.

9.2 Study 1 – Systematic Review: Discussion

9.2.1 Study Overview

The systematic review identified the available assessment tools which could be used by teachers within special education settings to measure adaptive behaviour, behaviour that challenges or autism-related behaviour of children with intellectual disabilities. A number of factors were then considered to determine which, if any, of the 26 assessments identified in 42 studies were appropriate for assessing pupils on the autism spectrum with intellectual disabilities. Relevant factors included the availability of the assessment, accessibility and ease of use by teachers, whether the assessments had been evaluated with a relevant population and with teacher respondents, and the outcome of the evaluation in relation to the measurement properties of the assessments. A number of assessments such as the TAPS and the ATEC had potential for the purposes outlined in this research, however still had notable limitations that would need to be addressed to ensure that the assessments were appropriate. The systematic review also highlighted some key issues relating to autism assessment in education which had implications for both the subsequent assessment development studies as

well as implications for educational practice. These will be discussed in more detail in the following sections.

9.2.2 Implications for the ABLE-Autism

9.2.2.1 Rationale

In the development of any assessment, it is first necessary to determine whether assessments which are appropriate for the relevant purposes already exist. Streiner and Kottner (2014) recognised that a vast number of scales which have been developed for a variety of purposes have never been used. This was evidenced by the recent systematic review by Provenzani et al. (2020) which looked at autism-specific outcome assessments and found that 69% of measures were only used within the literature once and only seven of the assessments were used in over 5% of the studies. A rationale of why a new assessment is needed, therefore, is of central importance in the early part of the assessment development process to ensure that a new assessment is not developed unnecessarily. The systematic review, conducted as Study 1 in this research, evidenced the research gap by showing the lack of robust, autism-specific assessments developed for use in schools and with input from teachers that have been evaluated in appropriate contexts, particularly within the United Kingdom.

9.2.2.2 Domains

Although the construct was defined gradually during the research process, the areas which were considered within the systematic review search formed the beginning of the barriers to learning construct definition and comprised the deductive part of this research. As the assessment was to be autism-specific and purposely developed to be relevant to the group of pupils with coexisting intellectual disabilities, it was necessary to consider the areas recognised in the literature that may have specific implications for this population. The specific educational needs of this group were outlined in Chapter 1 and many of these areas were recognised in the autism diagnostic criteria. The ICD-11 and DSM-V also acknowledged the different manifestations of these characteristics for individuals with intellectual disabilities as opposed to autistic individuals without coexisting conditions (APA, 2013; WHO, 2020). As a result of early informal scoping reviews, the categories of skills and behaviours were condensed as described in Chapters 3 and 4. However, key words such as *sensory*, *communication*, *functioning* and *emotional* were still included in the search to ensure that measures assessing these areas would be found.

9.2.3 Practice Implications

9.2.3.1 Lack of Educational Assessments in Peer-Reviewed Literature

The implications of the systematic review findings on the subsequent studies and remainder of the research are clear; the lack of current existing assessments for the specified purpose was evident and this was followed by the development of the ABLE-Autism. However, there are further implications of the systematic review findings on assessment in education more generally. The first is that assessments commonly used in education are not often found and evaluated in peer-reviewed literature. The reasons for this are unclear, although a contributing factor may be due to the divide in education and psychology as described in Chapter 2. My training and professional experience as a special needs teacher, for example, included accuracy and moderation with regard to assessment but, until my master's and doctoral studies, did not extend to measurement properties or validation. The lack of parity between the fields of psychology and education discussed in Chapter 2, particularly in terms of assessment, may be most obvious at classroom level when teachers deal with practicalities which feel removed from the theoretical aspects of assessment. In order to bridge this divide, a main focus throughout this thesis was on the usefulness of the assessment for teachers. By ensuring that the value teachers placed on the assessment was given as much weight as the sufficient measurement properties of the assessment, the importance of perspectives on both sides of the divide was recognised and balanced.

9.2.3.2 Inclusion of Teachers in Assessment Development

This leads on to perhaps the most important implication of the systematic review findings; that teachers are not often included in assessment development or evaluation. In specialised fields such as medicine, speech and language therapy and occupational therapy, professionals are usually included in the development of assessments in meaningful ways. However, education is policy-driven and subject to regular political reform compared to many other fields (Everton et al., 2007). Education is therefore not 'teacher-led' in the way that clinical fields may be led by relevant professionals. The findings of the systematic review suggested that the involvement of teachers in assessment is usually limited to that of respondents at the evaluation stage rather than providing meaningful contribution in the development of the assessment content. Including teachers in the various aspects of the assessment development process may not only ensure that they are acknowledged as experts and help them feel valued and empowered, but may also help to further bridge the divide

between psychology and education discussed above. As well as improving the reliability and validity of assessments, the more teachers are involved in development and validation processes the more they will understand the 'back end' of assessment and this may lead to improved assessment practice by teachers in classrooms and schools.

9.2.3.3 Opportunity for Autism-Specific Assessment in Special Schools

A further implication of the systematic review on assessment practices in special schools in the United Kingdom relates to the current assessment situation outlined in Chapter 2. As previously mentioned, the policy-driven nature of education impacts upon assessment development and assessment practices in schools. With focus heavily upon statutory assessment in previous years, recent changes to assessment policy have allowed schools and teachers more freedom to utilise assessment practice which best suits the needs of the school, teacher and pupils. Although the systematic review did identify autism-specific assessments, few of these were specific to autistic individuals with coexisting intellectual disabilities. The use of autism-specific assessments may be valuable to special needs teachers at a time where the advantages of assessment which is tailored to the needs of individual pupils or groups of pupils has been being recognised (McIntosh, 2015). The development of assessments created specifically for use with particular pupils with specific needs is in line with the social model of disability mentioned in Chapter 1. Rather than assess pupils using assessments which are normed on the general population, potentially resulting in floor effects, or considering these pupils to be "untestable" (Wolf-Schein, 1998, p.36), autism-specific assessments are an example of the adaptations that can be made in order to ensure that the abilities and progress of these pupils can be effectively identified and recorded. By developing assessments which have specific purposes relevant to groups of pupils or particular settings such as special schools, the most effective types and methods of assessment discussed in Chapter 2 can also be utilised. This will also allow the assessment to address some of the difficulties associated with assessing a particular group as described in Chapter 2. Similarly, education practices with this group of pupils may vary in different countries and, therefore, it is also necessary for assessments to be evaluated within the relevant education system in order to determine their real, practical value within an education system at a specific time.

9.3 Study 2a – Focus Groups: Discussion

9.3.1 Study Overview

There were three main research aims of the focus groups. Firstly, the focus groups aimed for special needs teachers to identify important areas of progress and barriers to learning for pupils on the autism spectrum in special schools to further refine the assessment construct and content. Secondly, the study looked to identify broad areas and specific skills or behaviours which the special needs teachers thought are important to assess for autistic pupils. Finally, the research aimed to consider the features of assessment tools which are useful to special needs teachers in the classroom. Analysis of the data from the two focus groups resulted in six main themes: *Autism-related barriers and atypical skill development, Overcoming barriers, Priorities for autistic pupils, 'True mastery', Assessing the bigger picture* and *Practicalities of assessment*. The implications of the information provided by teachers in the focus group discussions will be outlined in the following sections.

9.3.2 Implications for the ABLE-Autism

9.3.2.1 Construct Definition and Assessment Content

The results of the focus groups refined the definition of the barriers to learning construct and provided both broad areas and specific skills and behaviours to include in the content of the assessment. The focus group discussion acknowledged and provided support for the fact that the areas identified in Chapter 1, such as RRBs, functional communication and sensory needs, are considered barriers to learning for pupils on the autism spectrum by special school teachers. Teacher participants also indirectly acknowledged the uneven cognitive profile discussed in Chapter 1 by talking about non-linear progress of autistic pupils and the spiky profile. As a result of the focus group discussions, adaptive behaviour was not included as a subscale alone. Some aspects of adaptive behaviour, such as functional communication and self-regulation, were included as individual subscales. Independence and generalisation of skills, which teachers spoke of as priorities for their pupils with autism, were aspects of adaptive behaviour which were included throughout the discussions in relation to all skills and behaviours which were mentioned. It was therefore felt that adaptive behaviour could be addressed throughout the assessment and within individual items without the need for a specific and individual subscale.

An interesting outcome of the focus groups was that the teachers did not refer to 'behaviour that challenges' directly as a barrier to learning, but talked about the areas which cause or lead to behaviour that challenges such as anxieties and difficulties regulating emotion. This led to the subscale which was named *Emotion and Behaviour Regulation Barriers*. This subscale included items which describe skill gaps and behaviours which may lead to behaviour that is considered challenging (e.g. *EBRB2b. [Pupil] will respond appropriately to 'stop' and 'no'*) but also addressed the ability to self-regulate emotion and behaviour (e.g. *EBRB9. [Pupil] is able to stop or reduce inappropriate behaviour with support from adults*) as well as the ability to appropriately express their emotions and needs (e.g. *EBRB5c. [Pupil] will show that they want something to stop in an appropriate way*). Similarly, the subscale *Learning Behaviour Barriers* was included in the assessment as a result of the focus groups. Teachers talked about focus, attention, engagement and prerequisite learning skills which are often expected but often not explicitly taught or assessed. Many of these areas link to adaptive behaviour as well as a number of theories of autism described in Chapter 1 such as executive function and monotropism.

Teacher participants also reiterated that many of the areas discussed in Chapter 1, and subsequently included in the assessment, are linked to future outcomes and quality of life for the individual and their family. Teachers spoke passionately about collaborating with parents and supporting pupils and their families to be able to engage in various aspects of the community and preparing the pupils for life after school.

9.3.2.2 Assessment Features

During the focus groups, teachers were asked about current assessment practice in special schools. Themes in the discussions covered assessment practices more broadly in relation to autistic pupils as well as considering specific detail of assessment. The information provided by teachers on various features of assessment were, as far as possible, incorporated into the ABLE-Autism. Specific features of the ABLE-Autism were discussed in Chapter 7 but included positively worded skills and behaviours, spaces for teachers to provide descriptive comments, personalisation of items and the ability to account for regression.

A number of the themes and subthemes in the focus group discussions related to the difficulties assessing autistic pupils with coexisting intellectual disabilities which were discussed in Chapter 2. Teachers provided justification for autism-specific assessments by identifying that existing assessment frameworks may meet the needs of most pupils with

intellectual disabilities but were often inappropriate for pupils with autism. By developing assessments which have specific purposes relevant to particular groups of pupils or settings such as special schools, the most effective types and methods of assessment discussed in Chapter 2 can also be utilised. Avoiding linear assessment, for example, where skills are levelled or where certain items have to be completed before moving on, was important to account for the uneven profile and personal manifestation of skills and behaviour for individual pupils. A further example is the respondent report method used in the ABLE-Autism and the scale responses (*never* to *always*) which allow teachers to use their wider knowledge of the pupil rather than requiring direct observation or performance assessment. This addresses the requirement for assessment to allow for a holistic overview of a pupils' skills, abilities and behaviour, accounting for context and individual circumstances. Using assessment methods and features most suited to assessing this group of pupils will also result in the ability of the assessment to address some of the difficulties associated with assessing autistic pupils with intellectual disabilities which were described in Chapter 2.

9.3.3 Practice Implications

9.3.3.1 Teachers' Voices

The focus group study not only supported the early development of the ABLE-Autism but also addressed a wider research gap relating to special needs teachers and autistic pupils. This study was the only recent U.K.-based study which asked special needs teachers their opinions on teaching and assessment of pupils on the autism spectrum with intellectual disabilities. The lack of prior research asking opinions of special needs teachers was unexpected, given that they are the professionals who work most closely with the pupils and have a good understanding and specialist knowledge of their pupils' everyday learning needs. The results of the focus groups confirmed findings from more general previous research which suggested that special needs teachers' priorities are related to the specific needs and difficulties of these pupils rather than academic skills or more standard teaching and learning priorities of mainstream school teachers (Azad and Mandell, 2016; Helps et al., 1999).

Considering the recent assessment changes in England, there has also been a lack of previous research asking special needs teachers their opinions on assessment practice. In the focus groups, special needs teachers spoke about the fact that their autistic pupils do not fit the assessment systems and frameworks currently in use. This is an important finding with considerable implications for assessment practices in special schools. If assessment is to be

purposeful and meaningful rather than merely a 'tick-box exercise', it is important that feedback such as that provided in the focus groups is listened to and acted upon. Teachers did not merely speak of the ineffectiveness of current assessment, but also provided useful insight into what makes assessment useful. They discussed the practical side of assessment as well as discussing more abstract areas, for example, questioning the concept of 'true mastery'. Focus group conversations linked to the difficulties assessing pupils with autism and intellectual disabilities which were discussed in Chapter 2. Communication and motivation were addressed, for example, when teachers spoke of the need for strong relationships in order for the pupil to be able to really show "*what they're capable of*." The relevance and appropriateness of assessment content was also brought up by teachers, with recognition of the need to consider each pupil's abilities holistically and individually when assessing them. This linked to the discussion in Chapter 2 on measuring the things that are important rather than merely those that are easily measured (Resnick, 2017).

The focus group study showed that, although rarely asked, special needs teachers are willing to share their opinions on both the needs of their autistic pupils, their teaching priorities and their views on assessment. As has been acknowledged and evidenced throughout this thesis, teachers' input into assessment content and practice can have positive implications for both the content validity and development of new assessments as well as feedback on existing assessment systems and practices. In line with the social model of disability as well as principles and legislation around equality, diversity, inclusivity and non-discrimination, there is an additional argument that reasonable adjustments must be made to ensure that pupils on the autism spectrum are being assessed appropriately and this includes, for example, using assessments specifically developed for the needs of these pupils.

9.4 Study 2b – Delphi Exercise: Discussion

9.4.1 Study Overview

Following the focus group discussions, a list of items was developed as described in Chapter 7 and a modified Delphi exercise was conducted with special needs teachers to select items for inclusion in the ABLE-Autism. The aims of the research were to include, amend or reject potential items based on the ratings of relevance and comprehensibility of individual items as well as considering the comprehensiveness of the list of items as a whole (Terwee et al., 2018). A list of 86 items was considered by special needs teachers over two rounds. In the first round, 56 items met the criteria for inclusion and were retained. The remaining 30 items

were amended and, after three items were amalgamated with other items, 27 amended items were included in the second round. After the second round, 14 additional items met the threshold to be included in the final assessment leaving a total of 70 items for inclusion in the ABLE-Autism.

9.4.2 Implications for the ABLE-Autism

9.4.2.1 Relevance and Comprehensiveness

The high threshold for inclusion and the fact that a large number of items reached that threshold provided support that the items were appropriate and relevant to assess pupils on the autism spectrum in special schools. As many of the items reflected areas which were mentioned during the focus group discussion and also those discussed in Chapter 1, the results of the Delphi exercise added a quantitative element to the assessment development and triangulated the previous literature review and qualitative focus group data. Giving teachers the opportunity to rate items as 'not at all relevant' to pupils on the autism spectrum ensured that items weren't included merely because they covered skills and behaviours which could be easily measured (Resnick, 2017).

To ensure that the ABLE-Autism sufficiently covered all skills and behaviours relevant to barriers to learning for pupils on the autism spectrum in special schools, it was considered essential that special needs teachers were given an opportunity to suggest additional areas or items for inclusion in addition to rating individual items. A number of teachers made comments at the end of round one and suggested additional areas to consider. A number of these were incorporated into amended items which were included for teachers to rate in round two. One participant suggested a number of areas for inclusion related to a specific skills curriculum or behaviours and skills that were included elsewhere in the assessment. As these comments were from just one participant, they were not addressed at this point in the assessment development but additions to the assessment were to be considered by collecting further data from teachers who participated in the pilot testing of the assessment. The question on further areas for inclusion was not asked in the second round due to the fact that the list only represented the items which had not already been retained in round one. However, a comments box was provided for each item and at the end of the questionnaire and no teachers suggested that skills or behaviours were missing or that any further areas needed to be included. Comprehensiveness was revisited in the teacher feedback

during the evaluation stage, however, and will be discussed further in the discussion of the assessment evaluation study.

It is worth noting, as teachers were not asked whether skills and behaviours were *exclusively* relevant to their pupils on the autism spectrum, it is likely that some of the items included in the ABLE-Autism are also relevant to pupils with intellectual disabilities who are not autistic. This research has defended Cronbach's (1970) position that assessments should not be used for purposes beyond which they were developed and therefore only recommends use for individuals on the autism spectrum with intellectual disabilities. However, it is recognised that evaluating the assessment with different populations, such as those with intellectual disabilities who are not autistic or autistic individuals without intellectual disabilities, may provide further support for the need for autism-specific assessments. Furthermore, although there are clinical thresholds for intellectual disabilities (i.e. IQ < 70), it is understood that the distinction between those with and without intellectual disabilities is not absolute. Some pupils attending a mainstream school or without an official diagnosis of learning disabilities, for example, may have greater needs than some other individuals who have a diagnosis or attend a special school. It is recognised that the ABLE-Autism may be of use in some other limited contexts, however appropriate evaluation would need to first take place to determine suitability.

9.4.2.2 Comprehensibility

It was important for both the content validity and for the reliability of the ABLE-Autism that the assessment items were able to be understood by teachers. As with the question of item relevance, a high threshold for inclusion was considered appropriate as it was necessary to ensure that the items could be understood by any teachers who may use the assessment. Comprehensibility was also an important factor in the decision to include *any* qualified teachers in the Delphi exercise with no requirements of particular length of teaching experience. The assessment is intended to be able to be used by all teachers, including newly qualified teachers, for example. Therefore, only gauging the opinion of teachers with a specific number of years of teaching experience may have skewed the results and led to the inclusion of items which new teachers may not understand. A high threshold for inclusion (agreement by 80% of participants) was selected for similar reasons.

A number of items were amended after the first round of the Delphi exercise in response to participant comments. There were several comments included in the two rounds

of the Delphi exercise which related to the level of support the pupil required and the use of the word ‘appropriate’. This reflected the findings from the focus group where teachers mentioned precision as important in assessments. However, the level of support and a tighter description of what is considered ‘appropriate’ was not included in the items to guard against what teachers also described as assessments being too rigid and not flexible enough to reflect the needs of individual students and this was explained in Chapter 7. Instead, a feature included in the assessment was space for teachers to personalise items and provide comments if they required. This will be considered further in the discussion section on the assessment evaluation study.

9.4.3 Practice Implications

The Delphi study as part of this thesis had a number of implications beyond the selection of items for the ABLE-Autism. Firstly, the study outlined the use of this research method to allow teachers to contribute to assessment development and form a consensus on assessment content. As previous stages of the research have shown that teachers are not often included in assessment development, the Delphi exercise provides an example of a practical method to include teachers in educational assessment development. Additionally, the Delphi procedure used in this research provides an example of a high threshold for consensus using a four-point Likert scale along with use of the median in analysis of the data as suggested by Hsu and Sandford (2007). A final practical implication of this study includes the modifications made to the Delphi procedure. As mentioned in Chapter 3, the original Delphi method has been adapted and modified according to the research designs and aims across a range of disciplines (Hasson and Keeney, 2011). Modifications to the original Delphi method in the present research were outlined in Chapter 5. For example, the Delphi procedure in the present research included an amended second round. As consensus for exclusion was not as obvious using the four-point Likert scale, a dichotomous response was required from participants in the second round. This reflected the fact that each item was ultimately to be either included or excluded and this would be clearly communicated by participants in the second round with a dichotomous response option. Not only did this help indicate which items were not relevant or comprehensible, but may also have avoided participant burden and fatigue had participants been asked to complete the same questions on the same scale a second or even third time. Through this thesis, the use of the Delphi procedure for teachers to determine items for inclusion in an educational assessment has been outlined and may be used or further modified in similar studies in the future.

9.5 Study 3 – Assessment Evaluation: Discussion

9.5.1 Study Overview

The final stage of this research aimed to consider a number of measurement properties of the ABLE-Autism and provide preliminary evidence for the validity and reliability of the new measure. A total of 48 pupils on the autism spectrum who attend special schools were assessed at least once using the ABLE-Autism. Test-retest reliability was excellent for all subscales and the total score. A large effect size suggested that the ABLE-Autism is strongly correlated with the Teacher Autism Progress Scale, providing evidence for convergent validity. The ABLE-Autism subscales and scale as a whole showed a high degree of internal consistency. Teacher feedback was extremely positive which reflected the fact that teachers were consulted in all stages of the assessment development process.

9.5.2 Implications for the ABLE-Autism

9.5.2.1 Overview of Evaluated Measurement Properties

The evaluation of the ABLE-Autism was necessary to ensure that it was fit for purpose with the relevant population and respondents in an appropriate setting, the importance of which was outlined in Chapters 2 and 4. The evaluation study provided preliminary support for the reliability and validity of the ABLE-Autism, showing sufficient results over a number of measurement properties. In a practical sense, the results of the initial validity and reliability evaluations suggested that teachers were able to use the assessment consistently and accurately to measure relevant skills and behaviours. As some teachers assessed multiple pupils, multilevel modelling accounted for the fact that there may have been variation between the way different teachers assessed pupils.

9.5.2.2 Internal Consistency and Item Inclusion/Exclusion

The internal consistency evaluation was conducted to support future factor analysis studies. If later factor analysis suggests that the scale exhibits a different factor structure to the setup of the subscales, it may be that items are moved between subscales or that subscales are merged. As mentioned in Chapter 8, particularly high α values may suggest that some items are unnecessary or redundant as they measure the same construct. Educational assessment such as the ABLE-Autism, however, goes beyond the purpose of psychological assessment in merely measuring a construct. In assessments which consider skills, abilities and behaviour of pupils on the autism spectrum with intellectual disabilities, particular items

themselves may be important and relevant for teachers to assess. Haladyna and Roriguez (2013) discussed this as a "persistent conflict in item development" (p.15). They questioned:

> "Who decides the fate of a test item? In other words, do psychometric criteria trump all other factors?... Using solely psychometric criteria to decide the future of the item may improve reliability but at the expense of content-related validity evidence." (Haladyna and Rodriguez, 2013, p.15)

It may be that in a measure of depression, for example, one item might be unnecessary if it adequately indicates depression as well as another. However, in the current measure, although two items might be indicators of barriers to learning, it may be that teachers consider the skills represented by both of those items as relevant to their pupils and important to measure. Therefore, it would be important for changes in the items and structure of the ABLE-Autism due to psychometric evaluation to be supported by additional evaluation of content validity with teachers and other stakeholders.

9.5.2.3 Teacher-Teaching Assistant Reliability

Inter-rater reliability is an important measurement property and, particularly in education, is necessary to ensure that different teachers make the same judgements about the same pupil. Reliability of statutory educational assessment is often evaluated through teacher moderation, either within schools or between schools. When a judgement about a pupil is holistic, as teachers in the focus group discussions argued that assessment often is, it is difficult for this to be moderated without physical evidence such as video, photo or detailed narrative description. Teachers who participated in the focus group mentioned that they did not believe that this physical evidence was always necessary to the extent that it was sometimes required. Providing evidence for inter-rater reliability of an assessment would reduce the need for extensive moderation and the requirement for additional evidence of the judgements provided by respondents.

As mentioned in Chapter 8, inter-rater reliability was originally intended to be a part of the evaluation of the ABLE-Autism. The difficulty was finding two equally qualified respondents to complete the assessment who both had good knowledge of the pupils and their current abilities. It was recognised that TAs in special schools work closely with the pupils and usually have a good understanding of individual pupils' needs. Although TAs are unlikely to have the same understanding and skill as qualified class teachers, special schools

often require TAs to contribute to aspects of teaching and assessment in a way which TAs in mainstream schools do not. Although it would not be able to be referred to as inter-rater reliability, teacher-teaching assistant reliability was considered relevant and class TAs of pupils being assessed were invited to participate. Unfortunately, there were not enough data to analyse and report as only 14 pupils were assessed by four participating TAs. One TA completed the assessment for seven pupils, one further TA assessed three pupils and two TAs assessed two pupils each.

Informal exploration of comparisons between the teacher and TA data led to some interesting observations. For 10 of the 14 pupils assessed, the teacher assessments had a higher total score than the corresponding TA assessment, meaning that the teachers rated the pupil as having more barriers to learning than the TAs. There are a number of possible reasons why teachers may assess pupils as having more barriers to learning than TAs. This may be due to teachers noticing and identifying their pupils' needs and barriers as a result of their additional training. It may be that teachers hold higher expectations of pupils than TAs or perhaps that teachers have more experience of completing assessments and therefore TAs interpret the assessment items differently to teachers. The 14 TA-completed assessments contained 45 missing responses which equated over 4% of all responses. This was compared to only one missing response from the teacher responses relating to the same 14 pupils (< 0.01%). All but two of the missing responses were from one TA suggesting that the missing responses may be reflective of the assessment ability or experience of that particular TA. Due to the lack of data, these observations are merely speculative and should not be generalised. However, the data may indicate that teacher-TA reliability is something that requires further research and that the use of the ABLE-Autism should be evaluated further before it can be reliably and validly used by professionals other than the class teacher.

9.5.2.4 Responsiveness

As mentioned in Chapters 3 and 8, responsiveness is a key measurement property for an assessment which is designed to monitor change and progress. It is important that any changes in the assessment scores over time is reflective of actual change in the skills or behaviour rather than through error. The COSMIN manual, therefore, described responsiveness as referring to the "validity of the change score" (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018, p.60). Responsiveness can be evaluated by the testing of a hypothesis of expected change, for example through comparison with another

measure after intervention. The COSMIN manual explained that it is not enough for an assessment to merely measure change but it must also not over- or underestimate the change in the construct (Prinsen et al. 2018; Mokkink et al. 2018; Terwee et al. 2018). Although the responsiveness evaluation ultimately could not be conducted due to the COVID-19 pandemic, evaluating the responsiveness of the ABLE-Autism in the current study is likely to have faced some challenges. Comparing change scores with other assessments presented a difficulty due to the lack of autism-specific assessments designed for appropriate purposes which measured a similar construct. Using TAPS change scores for comparison may also prove difficult. As the TAPS is an assessment which was designed to be sensitive to weekly changes and ABLE-Autism scores are intended to be stable across a one to two-week time period, comparisons between the two instruments may be limited to direction of change scores. This could be triangulated by asking the teacher respondents about the expected direction of change in barriers to learning for their pupils prior to the third completion of the ABLE-Autism (Coster, 2013). Similarly, it would be important to gain teacher feedback on whether the ABLE-Autism effectively supports them to identify and record progress or regression, as meaningful change for some pupils may be much smaller than others due to their individual needs. Although potentially challenging, the evaluation of the responsiveness of the ABLE-Autism is a necessary next step in the validation of the assessment for the purposes of identifying progress for autistic pupils in special schools and is a recommendation for future research.

9.5.2.5 Teacher Feedback Overview

Arguably the most important part of the assessment evaluation is determining that it is useful for the purposes for which it was developed. In order to ensure that the ABLE-Autism was a valuable addition to assessments already available, feedback was sought from teachers who pilot tested the assessment to support the assessment development and content validity. The teacher feedback on the three areas of content validity was extremely positive. On a ten-point scale (where 10 was high and 1 was low) only one teacher rated any question below a 5. Mean scores for all four questions were above 8 and the median response for all questions was 8 or 9. Positive feedback was important at this stage of the research as it reflected the opinions of teachers who had used the assessment with their pupils rather than those rating assessment items in a hypothetical and more abstract manner.

9.5.2.6 Qualitative Feedback

The qualitative nature of the teacher comments in the feedback questionnaire was useful as it enabled teachers to explain their scores and provided a deeper insight into the teachers' thoughts on the assessment. There were 28 individual comments which were positive and supported the quantitative scores from the feedback questions. Teachers' descriptions of the ABLE-Autism included "*extremely useful*," "*straightforward*," "*relevant*," "*very clear*" and "*amazing assessment*." One teacher described interventions and environmental adaptations they had made as a result of using the assessment, providing evidence of the practical uses and benefits of the ABLE-Autism. Another teacher commented that they would recommend the assessment to colleagues and an additional teacher observed that the assessment is in line with a new, personalised curriculum their school is moving towards using. One teacher professed to "*love this assessment*" and identified that they were already seeing areas of improvement for their pupils.

Additional comments covered potential uses for the ABLE-Autism. Four individual comments mentioned that the assessment would be useful when a pupil was new to a class or school in order to identify their main barriers to learning. Four individual comments mentioned parents, with comments suggesting that it might be useful for parents to either complete the assessment or for the results to be shared with parents to support target setting and personalised learning. This is in line with Theme 2 from the focus group discussions in which teachers identified the importance of links with pupils' parents and the way this can support pupils to overcome their barriers to learning. Parent input and feedback into the assessment was considered in the research design and it is recognised that involvement by parents and families would be a valuable aspect of future research.

9.5.2.7 Feedback for Future Changes and Improvements

In relation to changes to the assessment in future, constructive feedback included comments on the features of the assessment, some suggestions for additional areas to include, and some queries around interpretation. Two teachers suggested that including an N/A box for items might be useful. This is a consideration for the future form of the assessment although this may reflect the fact that there is an element of artificiality to the assessment use during the pilot testing. This specific need may be resolved, for example, with the fact that teachers would not need to complete the whole assessment under ordinary circumstances if they did not consider it appropriate to do so and, instead, could choose to individually use

subscales which were most relevant to their pupils. Additionally, one teacher queried the use of the word 'appropriate' as some teachers did in the Delphi exercise. 'Appropriate' for the ABLE-Autism was intended to be interpreted by the teacher in relation to the expectations of the individual pupil in line with the intent for the assessment to be able to be individualised as far as possible for each pupil. It may be helpful to include this guidance for the term 'appropriate' in the instructions in future versions of the assessment and give examples, such as those which were outlined in Table 7 and discussed in Chapter 7. One teacher suggested that a lot of the skills wouldn't be possible for pupils with lower levels of cognition and another suggested that non-ambulant pupils would not be able to achieve some of the items relating to transition. The relevant items could have been personalised to the levels and expectations of each pupil but it may be that this wasn't made clear to the teachers using the assessment. Although the assessment was developed for pupils with moderate to severe learning difficulties, further evaluation comparing pupils of different abilities may be helpful in determining if the assessment is only useful for pupils who reach a particular level of cognition and understanding. Three comments mentioned the practical requirement of selecting a response option from a drop-down box as "*time consuming*" and something that could be improved. This feature was due to the format limitations in the assessment prototype and an easier way to select and deselect options would be included in future versions of the assessment.

When asked about the comprehensiveness of the assessment, suggestions included "*self-help skills, toileting, blowing nose, doing buttons, laces, zips, dressing,*" "*a little more about Emotional Regulation and sensory needs*" and "*more detailed smaller steps.*" These may all be appropriate skills and areas for addition to the ABLE-Autism. As these suggestions were from individual teachers, it is important that a wider number of teachers, parents or additional stakeholders are asked directly about these areas in future evaluations of the ABLE-Autism.

Only two of the 16 teachers who responded to the feedback questions made comments that the ABLE-Autism did not add to their current assessment practice. These two teachers identified other comprehensive assessments that they used (Routes for Learning and the SELLS curriculum). It is clear, as with any assessment, that not all teachers will find the assessment useful for all pupils at any given time. Although teachers' feedback was overwhelmingly positive, it may be useful to collect more detailed data on how teachers use the assessment as well as feedback from teachers during further evaluation, perhaps at

individual item level, to consider whether any further amendments can be made to ensure that the assessment is as useful as possible for the teachers who use it and the pupils who are being assessed.

Six of the items were marked for further consideration in the assessment evaluation as, although they reached the threshold for inclusion in the Delphi exercise, they received one score of 4 for comprehensibility in round one, meaning one teacher rated them not at all easy to understand (Appendix 5.3). None of the data related to these items appeared incongruent with any of the other items in the results of the evaluation study except for one of the items (*SB1*) which had three missing responses (two responses missing at Time 1 and one response missing at Time 2). This may have been due to difficulty understanding the item and therefore this item should also be addressed with teachers or parents in future research related to the ABLE-Autism to ensure it is able to be understood.

9.5.2.8 Use of the Personalisation Feature

The personalisation feature of the assessment was introduced in order to address the focus group and Delphi study comments around the rigidity of assessments and the fact that they do not account for the ways that skills and behaviour may manifest very differently for individual pupils when it comes to a heterogeneous condition such as autism. Theme 5 arising from the focus group discussions considered the necessity for assessments to address the range of needs of these "*unique and individual*" pupils. The importance of context in assessment was also included in this theme and the teachers acknowledged that wider context of both the pupil's abilities and the assessed skills and behaviours were not often addressed by assessments. Theme 6 incorporated the idea that pupils on the autism spectrum often did not 'fit' into assessment systems or frameworks due to the heterogeneity of the condition. In the Delphi exercise, teacher comments included questioning the meaning of 'appropriate' to describe skills and behaviours. This reflected the dichotomy mentioned in Chapter 5; not only is the meaning of 'appropriate' when relating to a skill or behaviour likely to differ between children but different teachers will also have varying interpretations of what is 'appropriate'.

In an attempt to address these issues, a personalisation feature was added to the ABLE-Autism. Each item had an optional 'What that looks like for this pupil' box where respondents had the option to include minor amendments to the items relating to context or personal targets. This box could also be used to specify if a pupil was working at a particular level of support, for example. During the evaluation of the ABLE-Autism, the effect of any

personalisation on the reliability and validity of items was going to be considered by comparing the reliability and validity of the items which had been personalised to the items where the feature had not been used and to the scale as a whole. As mentioned in Chapter 8, this feature was not used by any participants who tested the assessment so this evaluation could not be conducted.

There are a number of possible reasons why the personalisation feature was not used. Firstly, this may be such a new and unique feature for a school assessment that teachers may not have understood its purpose or potential uses. The assessment instructions contained a sentence describing how this feature could be used, however it may be that further information or examples of the use of this feature are necessary in order for teachers to understand its purpose. This was supported by a comment in the teacher feedback questionnaire which stated "*although the ability to add notes is helpful, with two places for comments/explaining what the skill looks like for the child, it gets a bit confusing.*" Secondly, as noted in Chapter 8, teachers were using the assessment somewhat artificially in the evaluation study and, therefore, the teachers may not have had the desire to explore this additional feature or may not have deemed it necessary for the pupils they were assessing. There is also the possibility that a personalisation feature such as this one is not useful for teachers. As the feature is optional, it will be left included in the ABLE-Autism for further consideration and analysis in future research. It is likely that teacher respondents will require specific instruction and examples of how this feature can be used. It may also be beneficial to further collaborate and discuss with teachers specifically on the question of how aspects of personalisation can be incorporated into the ABLE-Autism. Evaluation of how personalisation may affect the reliability and validity of the assessment will also be recommended.

9.5.2.9 Conclusions on Usability

It was understood that one assessment will not cover all of the potential barriers to learning that each individual pupil may encounter. However, the purpose of the ABLE-Autism was for the skills and behaviours of a particular group of pupils with specific needs to be assessed in a meaningful and effective way by special needs teachers. The comment in the feedback questionnaire by one teacher illustrated the ambition of the ABLE-Autism: "*It was like it was written for our class especially as we have a range of skills, abilities and behaviours.*" When an assessment has been developed specifically for pupils on the autism

spectrum with intellectual disabilities, with content formed and chosen by special needs teachers, the assessment is likely to be useful and valuable for those purposes. This simple comment reflected that the ABLE-Autism was, in fact, written for those pupils in that class and this was recognised by the teachers in their evaluation of the assessment.

9.5.3 Practice Implications

Practice implications of the evaluation of the ABLE-Autism comprise of those described above and result in the availability of an autism-specific school assessment which is both robust and useful to teachers. Further practice implications reflect those of the focus group and Delphi studies; identifying ways that teachers can be included in assessment research and the development of assessment tools. The systematic review in Stage 1 of this research provided justification for the need for assessments developed in collaboration with teachers and the evaluation study at Stage 3 provided evidence of the value of teacher input detailed in Stage 2. The ABLE-Autism fills a gap in the assessments available for teachers of pupils with autism and intellectual disabilities in special schools. It specifically addresses skills and behaviours that teachers wish to assess for these pupils and has proven to be useful to special needs teachers for the relevant purposes in the first pilot study. The ABLE-Autism is not related to a specific curriculum and its items do not follow a linear developmental trajectory. It can be used to supplement information about a pupil's difficulties and progress when teachers have a requirement for it, either alongside or in lieu of other curriculum based or school-wide assessments. The ABLE-Autism can be used alongside statutory assessments and allows teachers, schools and parents to identify important and relevant skills and behaviours the pupils have achieved, even when they may not appear to be accessing curriculum content. Appropriately identifying the barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities allows teachers to plan specific, individualised interventions and teaching to support pupils to gain skills which then enable them to access greater learning opportunities. Recording and monitoring progress or changes in learning barriers is also important for planning as well as to celebrate successes with the pupil, parents and school. Ultimately, reducing barriers and accessing and engaging in learning opportunities will result in the attainment of new, functional skills and improvements in outcomes and quality of life.

9.5.4 COVID-19

As the final stage of this research was conducted during early 2020, it is necessary to consider the impact of the global Covid-19 pandemic that occurred at this time. As discussed above, the initial aims of the research included a focus on responsiveness as a crucial measurement property and this required a third completion of the ABLE-Autism from teacher participants approximately six to eight weeks after the initial completion. Two teachers completed the assessment a third time in the spring term prior to the school closures however, as the remainder of teachers could not conduct the third completion, there were not enough data for analysis. Future evaluation of the responsiveness of the ABLE-Autism is one of the main recommendations of this research, due to the importance of this measurement property in ensuring that changes in barriers to learning and pupils' progress can be accurately identified and recorded.

COVID-19 in the United Kingdom also impacted upon participant recruitment; the author was in the process of liaising with a number of potential participants at the point of school closures and a number of further teachers who had signed the consent forms and received parent consent forms for their pupils then could not or chose not to participate. The number of second completions may also have been affected as, for a number of participants, this point in the data collection coincided with the beginning of the U.K. lockdown restrictions. Similarly, participants were contacted by email to complete the teacher feedback questionnaire during the time that school closures were in place and, although teachers were still working and had access to their emails, the number of responses may, again, have been affected.

It may also be useful to consider the uses and purpose of the ABLE-Autism in the context of the current global pandemic. Assessing progress in skills outside of the curriculum is important for pupils on the autism spectrum in special schools and, in circumstances such as the COVID-19 pandemic when usual teaching and school interventions are not able to take place and with a reduced focus on the formal curriculum, the importance of assessing these kinds of skills may be even more relevant and valuable. Some pupils will have returned to school after an extended absence and there have been broad changes in the both the physical school environment and the daily structure. At this time, assessment may be crucial to assess whether pupils' barriers have increased and to support teachers in identifying a pupil's greatest needs and the kinds of interventions to prioritise. For some pupils, it may also be

important to assess a baseline at the point they return to school in order to show progress when teaching and interventions begin. It is likely that the important non- or pre-curriculum skills and behaviours identified by teachers in the assessment development become even more relevant in the current circumstances.

9.6 Strengths and Weaknesses

Limitations of the individual studies have been described in each of the study chapters, however, there are a number of key limitations of the research as a whole and it is important that these limitations are outlined in order for them to be addressed in future research.

One main limitation of the research is that all of the studies contained small samples. The samples used throughout this research were convenience samples and relied on teachers' interest and willingness to give their time to participate during the short time period that each study was conducted. Larger and more representative samples would have allowed for more generalised conclusions and, therefore, replication of these studies with larger samples and in different areas of England is recommended. It may be useful to replicate the focus group discussions, for example, to determine whether similar themes would arise from discussions with teachers from a number of different schools across the country. Although the evaluation study was conducted with teachers and pupils across 13 special schools in various areas of England and Wales, a larger sample would allow for further analysis of various measurement properties of the ABLE-Autism (e.g. factor analysis, use of IRT and G Theory analyses).

A further limitation of the studies relates to the additional participant information. Diagnostic assessments were not confirmed or carried out for the pupils who were assessed in Stage 3 of the research. Additionally, demographic information (for example age, gender, ethnicity, years of teaching experience) was not collected for participating teachers. Although these demographics were not directly relevant to the studies, and therefore were not collected in line with data minimisation principles, patterns of responding influenced by particular demographics may have been missed.

As has been discussed throughout this thesis, an additional major limitation was that a responsiveness evaluation was not able to be carried out. In order to ensure that the ABLE-Autism appropriately identifies and measures changes in barriers to learning for autistic

pupils, it is important that the responsiveness of the assessment is determined in future studies.

An additional limitation is that, although gaining the views of other stakeholders was considered, only teachers were consulted in the course of this research. Practically, time and resource limitations meant that multiple stakeholder groups could not all be included in this initial development and preliminary evaluation of the new measure and teachers were therefore chosen as initial participants due to the importance of ensuring that the assessment content was appropriate for a special school and educational context. Consulting families, other professionals and autistic individuals with intellectual disabilities themselves where possible is a necessary step when considering all of the relevant stakeholders in the development of a new assessment. Autistic self-advocacy, the neurodiversity rights movement and autistic participation in research has grown in recognition and importance in recent decades (Crompton, 2020). Research has suggested that the views, opinions and priorities of different stakeholders in the field of autism, such as parent, professionals, researchers and autistic individuals themselves may not always align (Pellicano et al., 2014). The autism and neurodiversity 'nothing about us without us' movement aims to ensure participatory and inclusive research practice and ensure that autistic voices are heard (Crompton, 2020). Insights into autistic children's own priorities for their school life, as well as adults who have previously attended special schools, would be valuable in ongoing evaluations of the assessment to ensure that the difficulties and challenges included in the assessment align with autistic individuals' own present and long-term priorities. Although it may be difficult to gauge the views of the pupils for whom the assessment was developed due to complex communication difficulties, it must not be assumed that it would be impossible. It is crucial, as discussed in Chapter 1, that efforts are made for underrepresented groups such as those on the autism spectrum with moderate to severe intellectual disabilities to be included in research (Steadman et al. 2019). As Fletcher-Watson et al. (2019) recognised, "participatory research is both morally right and practically beneficial" (p.948) but should be relevant and specific to the topic to ensure that it is more than just a tokenistic gesture or a box-ticking exercise. It is also important to recognise the diverse experiences and expertise of the different stakeholder groups in order to work collaboratively towards achieving the ultimate aims of improving the lives and outcomes of autistic individuals. In addition to the inclusion of autistic children and adults' evaluation of the ABLE-Autism, teachers, for example, are likely to view the assessment in light of their professional knowledge and

experience of the education system and curriculum while parents may consider the skill areas in terms of those which might benefit family life as a whole. It is therefore important to also include the views of parents and families of autistic children with severe intellectual disabilities to ensure that the barriers included in the assessment are not only relevant to school but also to the everyday life of the child and their families. Additionally, some stakeholders may also belong to multiple groups, for example, teachers also being family members or parents also being autistic themselves, and these individuals may also have valuable insight. It is necessary in future to determine any differences in perspective from the stakeholder groups in order to ensure that any concerns or omissions relating to the ABLE-Autism are addressed. The involvement of these additional stakeholders would ensure that the assessment content adequately covers important quality of life and family outcomes as well as areas leading to good outcomes in adulthood and future life opportunities. It is recommended that these other stakeholders are included in the ongoing evaluation of the assessment to thoroughly and meaningfully evaluate the assessment content as well as the practical usefulness of ABLE-Autism.

Finally, a limitation of this research is that it has not considered how teachers who use the ABLE-Autism could go on to address the barriers to learning that the assessment has identified. The purpose of the ABLE-Autism is to identify which skills are present or absent rather than identify the causes of skill gaps or to suggest specific interventions and teaching strategies to use. The ABLE-Autism is limited in the fact that, in itself, the assessment does not provide further information on how the barriers identified can then be addressed. After identifying a pupil's barriers to learning, teachers may wish to use the information derived from the assessment to develop a description of the learning needs of their pupils which would directly inform the use of a specific curriculum, the use of certain interventions or further adaptations to the environment to support the pupil in learning and progressing in the relevant skills. Future research could help to determine which interventions, curricula or environmental adaptions teachers use as a result of the information that the ABLE-Autism provides. The assessment may also be used to show changes in skills and progress based on these adaptions to show which are most useful and effective in overcoming certain barriers. Additionally, although the focus of the ABLE-Autism is on skills and behaviours that can be taught and learnt by the pupil, it may be that specific additional assessments could be used alongside the ABLE-Autism. For example, a high score in the Sensory Barriers subscale may indicate that further in-depth assessment of a pupil's sensory profile is required along with

input from an occupational therapist specialising in sensory integration therapies. Similarly, if a number of barriers are shown to increase after a change of physical classroom environment or with a change of teacher, further evaluation of the classroom environment or teaching may be required to identify and evaluate whether additional environmental adaptations are necessary or if the interventions and teaching strategies are appropriate. Although it may be useful for teachers to be signposted to the appropriate next steps in order to address the causes of the barriers to learning based on the information that the ABLE-Autism provides, it is also important to ensure that the assessment information is not used in a rigid and prescriptive way. The flexibility to use the ABLE-Autism in conjunction with other assessments, curricula and interventions may also be a strength of the assessment. Holistic, individualised support specific to an individual's unique needs along with the aim of overcoming barriers to learning which will result in opportunities and increased independence, regardless of current level of ability, is also in line with the discussion on the social model of disability in Chapter 1.

Despite these limitations, this research has explored and addressed areas with a notable lack of prior research. Little, if any, previous research had focused on the assessment of pupils on the autism spectrum with intellectual disabilities in special schools in England and special needs teachers have been an underused resource in the selection of content and development of assessments that can be used in schools. The development of the ABLE-Autism has resulted in an addition to the available autism-specific assessments which can be used in special schools. Furthermore, the modifications made to research methods (e.g. the Delphi exercise) for these purposes have resulted in practical ways of further including special needs teachers in assessment research. As was discussed in Chapter 3, the inclusion of both qualitative and quantitative aspects of this mixed methods research accounted for the weaknesses in the two different approaches. Conducting the research from an epistemological position of pragmatism ensured that the research methods were selected due to their suitability to answer the research questions (Johnson and Onwuegbuzie, 2004). The quantitative analysis in the Delphi exercise balanced the qualitative aspects of the focus group study when developing the assessment content and a reflective journal was used to guard against subjective bias in the facilitation of the focus groups and the analysis of the data. Stage 3, being largely quantitative, ensured objective evaluation of the ABLE-Autism, yet qualitative elements such as teacher comments in the assessment feedback questionnaire were included to give deeper meaning and reasoning to the feedback scores. In this way, the

pragmatic, mixed methods research design was an inherent strength of the project through the differences of qualitative and quantitative research each accounting for some of the limitations of the other (Hammond, 2005).

9.7 Conclusions and Recommendations for Future Research

The research aim of this thesis was to consider how the barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities could be effectively identified and monitored through robust assessment by teachers in special schools. In the literature review, the unique characteristics and needs of individuals on the autism spectrum with coexisting intellectual disabilities were explored. In Chapter 1, the ways that some of these needs and characteristics could create barriers to learning for these pupils were discussed. The links between the characteristics and needs of this particular group were considered with reference to some of the main theories of autism. The difficulties in assessing this population in special schools were then addressed in Chapter 2. This included an outline of the current assessment practice and policy in special schools, the purposes and types of assessment in special education and the requirements of robust and effective assessment. The findings of the systematic review in Stage 1 of the research highlighted the research gap by identifying the lack of autism-specific assessments which could be used by teachers in special schools and which had been appropriately evaluated for the relevant purposes. The systematic review results justified the development of a new assessment which was autism-specific, developed in collaboration with teachers and covered areas which teachers identified as barriers to learning for these pupils in an education context. In Stage 2 of the research, the results of the focus groups further defined the 'barriers to learning' construct and specified both broad areas and specific skills and behaviours to be included in the assessment. The teacher focus groups also enabled teachers to discuss problems with special school assessments and how these might be overcome as well as the aspects of assessment they find useful. The Delphi exercise in the second part of Stage 2 resulted in teachers selecting 70 items to be included in the new assessment, the ABLE-Autism. The pilot test and evaluation of the ABLE-Autism in Stage 3 provided initial evidence for reliability and validity of the new assessment. This stage of the research also resulted in positive teacher feedback and supported the idea that teacher assessments are likely to have more practical benefits for use in the classroom if they are developed alongside the teachers who will use them.

The findings of this thesis have led to a number of recommendations for further research in the future, both to adapt and improve the ABLE-Autism and also in respect of wider special school and autism-based research practice. Firstly, as assessment evaluation is an ongoing process, further evaluation of the ABLE-Autism is recommended. Measurement properties which have not yet been evaluated should be prioritised. Factor analysis should be conducted to determine the internal structure of the assessment, as well as inter-rater reliability evaluations to consider whether different raters interpret the items in a consistent manner. Responsiveness should be evaluated to establish whether the ABLE-Autism can adequately measure change in barriers to learning. Analyses related to IRT and G Theory should be considered for further item-level evaluation and to identify and distinguish sources of measurement error. Additionally, teachers should continue to provide feedback on the assessment, including whether it is useful and appropriate for its intended purpose and how it can be improved. Further instructions for and examples of the use of the personalisation feature should be given and further evaluations conducted in order to determine whether this feature affects validity and reliability of the items and the overall assessment. It is also important for the ABLE-Autism to be evaluated when it is used in a more natural and authentic context than the initial evaluation study, for example when teachers independently select and use the assessment with particular pupils due to their specific needs. It may also be interesting and appropriate for future research to determine the assessment norms with the relevant population or to compare the assessment results between subgroups of pupils (e.g. pupils of different ages, males and females). It may be appropriate in future to consider curricula, other assessments or interventions which may support teachers in addressing the barriers to learning identified by the ABLE-Autism. Finally, it is necessary for other stakeholders such as parents and pupils to provide feedback on the content of the ABLE-Autism to ensure that the assessment includes skills and behaviours which are not only important in the classroom, but also support individual and family quality of life.

This research has also resulted in recommendations for wider research and practice relating to assessment and the education of pupils on the autism spectrum with coexisting intellectual disabilities. The research supports conclusions drawn from commentators such as Stedman et al. (2019), Jordan (2001), Matson and Shoemaker (2009) and Hurley and Levitas (2007); that the interaction between autism and intellectual disabilities results in specific needs for this group of individuals which are different to those with either intellectual disabilities or autism alone. It is important, where appropriate, that these individuals are

included in research and considered as a distinct population in education. As has been the case with assessment, merely including these individuals within a wider population of intellectual disabilities may mean that important characteristics and needs specific to this group may not be recognised or met. Similarly, the increase in autism research has coincided in an underrepresentation of autistic participants with coexisting intellectual disabilities and it is important that this imbalance is addressed, particularly in areas such as education research. This research also highlights, not only appropriate methods, but also the benefits of including teachers in assessment development processes. The inclusion of teachers in the development of the ABLE-Autism is likely to be a contributing factor to the positive teacher feedback and promising preliminary validity and reliability evaluations. It is recommended that teachers are included in the development and evaluation of school and teacher assessments to ensure that the assessments are, not only robust in terms of their measurement properties, but ultimately useful for the purposes for which they were developed.

References

Abbott, C., & Lucey, H. (2005). Symbol communication in special schools in England: the current position and some key issues. *British Journal of Special Education, 32*(4), 196-201. https://doi.org/10.1111/j.1467-8578.2005.00397.x

Adams, D., MacDonald, L., & Keen, D. (2019). Teacher responses to anxiety-related behaviours in students on the autism spectrum. *Research in Developmental Disabilities, 86*, 11-19. https://doi.org/10.1016/j.ridd.2018.12.009

Adelman, H. S., & Taylor, L. (2006). *The school leader's guide to student learning supports: New directions for addressing barriers to learning*. Corwin Press.

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress, 40*(7), 64-65.

Aman, M. G., Tassé, M. J., Rojahn, J., & Hammer, D. (1996). The Nisonger CBRF: A child behavior rating form for children with developmental disabilities. *Research in Developmental Disabilities, 17*(1), 41-57. https://doi.org/10.1016/0891-4222(95)00039-9

Amaral, D. G. (2017). Examining the causes of autism. *Cerebrum*. Cer-01-17.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) American Psychiatric Publishing.

Amos, T., & Pearse, N. (2008). Pragmatic research design: An illustration of the use of the Delphi technique. *Electronic Journal of Business Research Methods, 6*(2), 95-102.

Amstad, M., & Müller, C. M. (2020). Students' problem behaviors as sources of teacher stress in special needs schools for individuals with intellectual disabilities. *Frontiers in Education, 4*(159), 1-11. https://doi.org/10.3389/feduc.2019.00159

Aricak, O. T., & Oakland, T. (2010). Multigroup confirmatory factor analysis for the teacher form, ages 5 to 21, of the adaptive behavior assessment system-II. *Journal of Psychoeducational Assessment, 28*(6), 578-584. https://doi.org/10.1177/0734282909350209

Arnold, S., & Reed, P. (2016). Reading assessments for students with ASD: A survey of summative reading assessments used in special educational schools in the UK. *British Journal of Special Education, 43*(2), 122-141. https://doi.org/10.1111/1467-8578.12127

Atilgan, H. (2013). Sample size for estimation of G and phi coefficients in Generalizability Theory. *Eurasian Journal of Educational Research*, *51*, 215-227.

Azad, G., & Mandell, D. S. (2016). Concerns of parents and teachers of children with autism in elementary school. *Autism*, *20*(4), 435-441. https://doi.org/10.1177/1362361315588199

Bailey Jr, D. B., & Wolery, M. (1989). *Assessing infants and preschoolers with handicaps*. Merrill Publishing Co.

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317-350. https://doi.org/10.1080/0969594X.2017.1319337

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *57*(12), 1424-1433. https://doi.org/10.1111/jcpp.12609

Bal, V. H., Kim, S. H., Cheong, D., & Lord, C. (2015). Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age. *Autism*, *19*(7), 774-784. https://doi.org/10.1177/1362361315575840

Bancroft, K., Batten, A., Lambert, S., & Madders, T. (2012) *The Way We Are: Autism in 2012*. The National Autistic Society.

Banerjee, R. (2016). Checklists and Rating Scales. In Couchenour, D., & Chrisman, J. K. (Eds.). *The SAGE Encyclopedia of Contemporary Early Childhood Education* (pp. 196-199). SAGE Publications.

Bangerter, A., Ness, S., Aman, M. G., Esbensen, A. J., Goodwin, M. S., Dawson, G., & Harris, A. (2017). Autism behavior inventory: A novel tool for assessing core and associated symptoms of autism spectrum disorder. *Journal of Child and Adolescent Psychopharmacology*, *27*(9), 814-822. https://doi.org/10.1089/cap.2017.0018

Barney, C. C., Andersen, R. D., Defrin, R., Genik, L. M., McGuire, B. E., & Symons, F. J. (2020). Challenges in pain assessment and management among individuals with intellectual and developmental disabilities. *Pain Reports*, *5*(4), 1-8. https://doi.org/10.1097/PR9.0000000000000822

Baron-Cohen, S. (2009). Autism: the empathizing-systemizing (ES) theory. *Annals of the New York Academy of Sciences*, *1156*(1), 68-80. https://doi.org/10.1111/j.1749-6632.2009.04467.x

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Baxter, A., Brugha, T., Erskine, H., Scheurer, R., Vos, T., & Scott, J. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine, 45*(3), 601-613. doi:10.1017/S003329171400172X

Beadle-Brown, J., Murphy, G., & DiTerlizzi, M. (2009). Quality of life for the Camberwell cohort. *Journal of Applied Research in Intellectual Disabilities*, *22*(4), 380-390. https://doi.org/10.1111/j.1468-3148.2008.00473.x

Beadle-Brown, J., Roberts, R., & Mills, R. (2009). Person-centred approaches to supporting children and adults with autism spectrum disorders. *Tizard Learning Disability Review, 14*(3), 18-26. https://doi.org/10.1108/13595474200900024

Beauchamp, T. L. (2007). The 'four principles' approach to health care ethics. In Ashcroft, R. E., Dawson, A., Draper, H., and McMillan, J. (Eds.) *Principles of Health Care Ethics (Second Edition)*, pp.3-10. John Wiley & Sons.

Berchtold, A. (2016). Test–retest: agreement or reliability? *Methodological Innovations*, *9*, 1-7. http://doi.org/10.1177/205979911667287

Berry, K., Russell, K., & Frost, K. (2018). Restricted and repetitive behaviors in autism spectrum disorder: A review of associated features and presentation across clinical populations. *Current Developmental Disorders Reports, 5*(2), 108-115. https://doi.org/10.1007/s40474-018-0139-0

Bertrand, J., Mars, A., Boyle, C., Bove, F. Yeargin-Allsop, M., & Decloufle, P. (2001). Prevalence of autism in a United States population: The Brick Township, New Jersey, investigation. *Pediatrics, 108*(5), 1155-1161. http://doi.org/10.1542/peds.108.5.1155

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*. https://doi.org/10.3389/fpubh.2018.00149

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, *48*(3), 587-605. https://doi.org/10.1111/j.1744-6570.1995.tb01772.x

Boucher, J. M. (2017). *Autism Spectrum Disorder: Characteristics, Causes and Practical Issues.* Sage.

Bowring, D. L., Totsika, V., Hastings, R. P., Toogood, S., & Griffith, G. M. (2017). Challenging behaviours in adults with an intellectual disability: A total population study and exploration of risk indices. *British Journal of Clinical Psychology*, *56*(1), 16-32. https://doi.org/10.1111/bjc.12118

Bradshaw, J., Gillespie, S., Klaiman, C., Klin, A., & Saulnier, C. (2019). Early emergence of discrepancy in adaptive behavior and cognitive skills in toddlers with autism spectrum disorder. *Autism*, *23*(6), 1485–1496. https://doi.org/10.1177/1362361318815662

Brett, D., Warnell, F., McConachie, H., & Parr, J. (2016). Factors affecting age at ASD diagnosis in UK: No Evidence that Diagnosis Age has Decreased Between 2004 and 2014. *Journal of Autism & Developmental Disorders*, *46*(6), 1974-1984. https://doi.org/10.1007/s10803-016-2716-6

Brooks, J., McCluskey, S., Turley, E., & King, N. (2015). The utility of template analysis in qualitative psychology research. *Qualitative Research in Psychology*, *12*(2), 202-222. https://doi.org/10.1080/14780887.2014.955224

Brown, G. T., & Abdulnabi, H. H. (2017). Evaluating the quality of higher education instructor-constructed Multiple-choice Tests: impact on student grades. *Frontiers in Education*, *2*, 24. https://doi.org/10.3389/feduc.2017.00024

Brown, E. C., Aman, M. G., & Havercamp, S. M. (2002). Factor analysis and norms for parent ratings on the Aberrant Behavior Checklist-Community for young people in special education. *Research in Developmental Disabilities*, *23*(1), 45-60. https://doi.org/10.1016/S0891-4222(01)00091-9

Brugha, T. S., Doos, L., Tempier, A., Einfeld, S., & Howlin, P. (2015). Outcome measures in intervention trials for adults with autism spectrum disorders; a systematic review of assessments of core autism features and associated emotional and behavioural problems.

International Journal of Methods in Psychiatric Research, 24(2), 99-115. https://doi.org/10.1002/mpr.1466

Brundrett, M., & Rhodes, C. (2013). Theories of educational research. In Brundrett, M., & Rhodes, C. *Researching Educational Leadership and Management* (pp.12-23). Sage Publications Ltd.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology*, *35*(1), 75-92. https://doi.org/10.2307/590553

Burack, J. A., & Volkmar, F. R. (1992). Development of low-and high-functioning autistic children. *Journal of Child Psychology and Psychiatry*, *33*(3), 607-616. https://doi.org/10.1111/j.1469-7610.1992.tb00894.x

Burgess, A., & Gutstein, S. (2007). Quality of life for people with autism: Raising the standard for evaluating successful outcomes. *Child and Adolescent Mental Health, 12*(2), 80-86. https://doi.org/10.1111/j.1475-3588.2006.00432.x

Carulla, L. S., Reed, G. M., Vaez-Azizi, L. M., Cooper, S. A., Leal, R. M., Bertelli, M., & Girimaji, S. C. (2011). Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry*, *10*(3), 175.

Case-Smith, J., Weaver, L. L., & Fristad, M. A. (2015). A systematic review of sensory processing interventions for children with autism spectrum disorders. *Autism*, *19*(2), 133-148. https://doi.org/10.1177/1362361313517762

Chakrabarti, S., & Fombonne, E. (2005). Pervasive developmental disorders in preschool children: confirmation of high prevalence. *American Journal of Psychiatry, 162*(6), 1133–1141. https://doi.org/10.1176/appi.ajp.162.6.1133

Charman, T., Howlin, P., Berry, B., & Prince, E. (2004). Measuring developmental progress of children with autism spectrum disorder on school entry using parent report. *Autism, 8*(1), 89-100. https://doi.org/10.1177/1362361304040641

Charman, T., Jones, C.R., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011a). Defining the cognitive phenotype of autism. *Brain Research, 1380*, 10-21. https://doi.org/10.1016/j.brainres.2010.10.075

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011b). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(3), 619-627. https://doi.org/10.1017/S0033291710000991

Chatham, C. H., Taylor, K. I., Charman, T., Liogier D'Ardhuy, X., Eule, E., Fedele, A., & San Jose Caceres, A. (2018). Adaptive behavior in autism: Minimal clinically important differences on the Vineland-II. *Autism Research*, *11*(2), 270-283. https://doi.org/10.1002/aur.1874

Cheminais, R. (2013). *How to Create the Inclusive Classroom: Removing Barriers to Learning*. Routledge.

Chiang, H. M. (2008). Expressive communication of children with autism: the use of challenging behaviour. *Journal of Intellectual Disability Research*, *52*(11), 966-972. https://doi.org/10.1111/j.1365-2788.2008.01042.x

Chiarotti, F., & Venerosi, A. (2020). Epidemiology of autism spectrum disorders: A review of worldwide prevalence estimates since 2014. *Brain Sciences*, *10*(5), 274. https://doi.org/10.3390/brainsci10050274

Chowdhury, M., Aman, M. G., Scahill, L., Swiezy, N., Arnold, L. E., Lecavalier, L., & Sukhodolsky, D. (2010). The Home Situations Questionnaire-PDD version: Factor structure and psychometric properties. *Journal of Intellectual Disability Research*, *54*(3), 281-291. https://doi.org/10.1111/j.1365-2788.2010.01259.x

Cluley, V. (2018). From “Learning disability to intellectual disability”—Perceptions of the increasing use of the term “intellectual disability” in learning disability policy, research and practice. *British Journal of Learning Disabilities*, *46*(1), 24-32. https://doi.org/10.1111/bld.12209

Cohen, I. L. (2003). Criterion-related validity of the PDD Behavior Inventory. *Journal of Autism and Developmental Disorders, 33*(1), 47-53. https://doi.org/10.1023/A:1022278420716

Cohen, I. L., Schmidt-Lackner, S., Romanczyk, R., & Sudhalter, V. (2003). The PDD Behavior Inventory: A rating scale for assessing response to intervention in children with pervasive developmental disorder. *Journal of Autism and Developmental Disorders*, *33*(1), 31-45. https://doi.org /10.1023/A:1022226403878

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250-1262. https://doi.org/10.1016/j.jbusres.2008.01.013

Coster, W. J. (2013). Making the best match: selecting outcome measures for clinical trials and outcome studies. *American Journal of Occupational Therapy*, *67*(2), 162-170. https://doi.org/10.5014/ajot.2013.006015

Coster, W. J., Mancini, M. C., & Ludlow, L. H. (1999). Factor structure of the school function assessment. *Educational and Psychological Measurement, 59*(4), 665-677. https://doi.org/10.1177/00131649921970099

Crompton, C. (2020) Autism, Neurodiversity, and the Humanities. In Fletcher-Watson, B. & Phillips, J., (Eds.), *Humanities of the Future: Perspectives from the Past and Present* (Volume 21) (pp. 59-63). IASH Occasional Papers.

Cronbach, L. J. (1970). *Essentials of Psychological Testing, Third Edition.* Harper and Row.

Crook, B., Tomlins, R., Bancroft, A., & Ogi, L. (2016). 'So often they do not get recruited': exploring service user and staff perspectives on participation in learning disability research and the barriers that inhibit it. *British Journal of Learning Disabilities*, *44*(2), 130-137. https://doi.org/10.1111/bld.12120

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458-467. https://doi.org/10.1287/mnsc.9.3.458

Dang, K., Bent, S., Lawton, B., Warren, T., Widjaja, F., McDonald, M. G., & Hendren, R. L. (2017). Integrating autism care through a school-based intervention model: A pilot study. *Journal of Clinical Medicine*, *6*(10), 97. https://doi.org/10.3390/jcm6100097

Davies, P. L., Soon, P. L., Young, M., & Clausen-Yamaki, A. (2004). Validity and reliability of the school function assessment in elementary school students with disabilities. *Physical and Occupational Therapy in Pediatrics*, *24*(3), 23-43. https://doi.org/10.1300/J006v24n03_03

Davis, A. (1998). *The Limits of Educational Assessment.* Blackwell.

Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of Delphi studies. *The Electronic Journal of Business Research Methodology*, *3*(2), 103-116.

De Bildt, A., Sytema, S., Kraijer, D., Sparrow, S., & Minderaa, R. (2005). Adaptive functioning and behaviour problems in relation to level of education in children and adolescents with intellectual disability. *Journal of Intellectual Disability Research, 49*(9), 672-681. https://doi.org/10.1111/j.1365-2788.2005.00711.x

Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy & Practice, 8*(2), 113-133. https://doi.org/10.1080/09695940123828

Department for Children, Schools and Families. (2014). *Designing for disabled children and children with special educational needs.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/276698/Building_Bulletin_102_designing_for_disabled_children_and_children_with_SEN.pdf [Accessed 22 November 2020]

Department for Education. (2014) *National Curriculum and Assessment from September 2014: Information for Schools.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/358070/NC_assessment_quals_factsheet_Sept_update.pdf [Accessed 18 February 2018].

Department for Education. (2019a) *Special Educational Needs in England: January, 2019.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/814244/SEN_2019_Text.docx.pdf [Accessed 14 November 2019]

Department for Education. (2019b) *Statement of SEN and EHC Plans: England, 2019.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/805014/SEN2_2019_text.pdf [Accessed 9 March 2020]

Detmar, S. B., Bruil, J., Ravens-Sieberer, U., Gosch, A., & Bisegger, C. (2006). The use of focus groups in the development of the KIDSCREEN HRQL questionnaire. *Quality of Life Research, 15*(8),1345-1353. https://doi.org/10.1007/s11136-006-0022-z

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203-1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Dover, C. J., & Le Couteur, A. (2007). How to diagnose autism. *Archives of Disease in Childhood, 92*(6), 540-545. https://doi.org/10.1136/adc.2005.086280

Dumas, H., Fragala-Pinkham, M., Haley, S., Coster, W., Kramer, J., Kao, Y. C., & Moed, R. (2010). Item bank development for a revised pediatric evaluation of disability inventory (PEDI). *Physical and Occupational Therapy in Pediatrics*, *30*(3), 168-184. https://doi.org/10.3109/01942631003640493

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, *14*(7), 1-11. https://doi.org/10.7275/jg4h-rb87

Dunn, K., Rydzewska, E., Macintyre, C., Rintoul, J., & Cooper, S. A. (2019). The prevalence and general health status of people with intellectual disabilities and autism co-occurring together: a total population study. *Journal of Intellectual Disability Research*, *63*(4), 277-285. https://doi.org/10.1111/jir.12573

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155. https://doi.org/10.1037/1082-989X.5.2.155

Einfeld, S. L., & Tonge, B. J. (1991). Psychometric and clinical assessment of psychopathology in developmentally disabled children. *Australia and New Zealand Journal of Developmental Disabilities*, *17*(2), 147-154. https://doi.org/10.1080/07263869100034361

Ellison, K. S., Bundy, M. B., Wygant, D. B., & Gore, J. S. (2016). Don't forget about the teachers! Parent and teacher BASC-2 reports of children on the DSM-5 Autism Spectrum. *Journal of Mental Health Research in Intellectual Disabilities*, *9*(3), 119-132. https://doi.org/10.1080/19315864.2016.1181812

Emerson, E. (2001). *Challenging Behaviour: Analysis and Intervention in People with Severe Intellectual Disabilities*. Cambridge University Press.

Emily, G., & Grace, I. (2015). Family quality of life and ASD: The role of child adaptive functioning and behavior problems. *Autism Research*, *8*(2), 199-213. https://doi.org/10.1002/aur.1442

Everton, T., Turner, P., Hargreaves, L., & Pell, T. (2007). Public perceptions of the teaching profession. *Research Papers in Education*, *22*(3), 247-265. https://doi.org/10.1080/02671520701497548

Fayers, P. M., & Hand, D. J. (2002). Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *165*(2), 233-253. https://doi.org/10.1111/1467-985X.02020

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., & Pellicano, E. (2019). Making the future together: Shaping autism research through meaningful participation. *Autism*, *23*(4), 943-953.

Fleuren, B. P., van Amelsvoort, L. G., Zijlstra, F. R., de Grip, A., & Kant, I. (2018). Handling the reflective-formative measurement conundrum: A practical illustration based on sustainable employability. *Journal of Clinical Epidemiology*, *103*, 71-81. https://doi.org/10.1016/j.jclinepi.2018.07.007

Flick, U., von Kardoff, E., & Steinke, I. (Eds.). (2004). *A Companion to Qualitative Research*. Sage.

Floyd, R. G., Shands, E. I., Alfonso, V. C., Phillips, J. F., Autry, B. K., Mosteller, J. A., & Irby, S. (2015). A systematic review and psychometric evaluation of adaptive behavior scales and recommendations for practice. *Journal of Applied School Psychology*, *31*(1), 83-113. https://doi.org/10.1080/15377903.2014.979384

Frith, U. (2003). *Autism Explaining the Enigma*. Blackwell.

Furr, M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. SAGE Publications Ltd.

Gallagher, D. J., Connor, D. J., & Ferri, B. A. (2014). Beyond the far too incessant schism: Special education and the social model of disability. *International Journal of Inclusive Education*, *18*(11), 1120-1142. https://doi.org/10.1080/13603116.2013.875599

Garg, A. X., Hackam, D., & Tonelli, M. (2008). Systematic review and meta-analysis: When one study is just not enough. *Clinical Journal of the American Society of Nephrology*, *3*(1), 253-260. https://doi.org/10.2215/CJN.01430307

Gargon, E., Crew, R., Burnside, G., & Williamson, P. R. (2019). Higher number of items associated with significantly lower response rates in COS Delphi surveys. *Journal of Clinical Epidemiology*, *108*, 110-120. https://doi.org/10.1016/j.jclinepi.2018.12.010

Georgiades, S., Szatmari, P., & Boyle, M. (2013). Importance of studying heterogeneity in autism. *Neuropsychiatry*, *3*, 123-125. https://doi.org/10.2217/npy.13.8

Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. Routledge.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist. 18*(8), 519-521. https://doi.org/10.1111/j.1745-3992.1994.tb00561.x

Gore, N. J., McGill, P., Toogood, S., Allen, D., Hughes, J. C., Baker, P., & Denne, L. D. (2013). Definition and scope for positive behavioural support. *International Journal of Positive Behavioural Support*, *3*(2), 14-23.

Green, R. A. (2014). The Delphi technique in educational research. *Sage Open*, *4*(2),1-8. https://doi.org/10.1177/2158244014529773

Grove, R., Baillie, A., Allison, C., Baron-Cohen, S., & Hoekstra, R. A. (2013). Empathizing, systemizing, and autistic traits: Latent structure in individuals with autism, their parents, and general population controls. *Journal of Abnormal Psychology*, *122*(2), 600-609. https://doi.org/10.1037/a0031919

Grzadzinski, R., Donovan, K., Truong, K., Nowell, S., Lee, H., Sideris, J., & Watson, L. R. (2020). Sensory Reactivity at 1 and 2 Years Old is Associated with ASD Severity During the Preschool Years. *Journal of Autism and Developmental Disorders*, *50*(11), 3895-3904. https://doi.org/10.1007/s10803-020-04432-4

Gunn, K. C., & Delafield-Butt, J. T. (2016). Teaching children with autism spectrum disorder with restricted interests: A review of evidence for best practice. *Review of Educational Research*, *86*(2), 408-430. https://doi.org/10.3102/0034654315604027

Gupta, A., & Singhal, N. (2009). Language and learning skills and symptoms in children with autistic spectrum disorders. *Asia Pacific Disability Rehabilitation Journal. 20*(2), 59–83.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and Validating Test Items*. Routledge.

Hammond, C. (2005). The wider benefits of adult learning: An illustration of the advantages of multi-method research. *International Journal of Social Research Methodology*, *8*(3), 239-255. https://doi.org/10.1080/13645570500155037

Hannes, K. (2011). Chapter 4: Critical appraisal of qualitative research. In: Noyes, J., Booth, A., Hannes, K., Harden, A., Harris, J., Lewin, S., & Lockwood, C. (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions.* Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group, 2011. http://cqrmg.cochrane.org/supplemental-handbook-guidance

Hanratty, J., Livingstone, N., Robalino, S., Terwee, C. B., Glod, M., Oono, I. P., & McConachie, H. (2015). Systematic review of the measurement properties of tools used to measure behaviour problems in young children with autism. *PLoS One, 10*(12), 1-21. https://doi.org/10.1371/journal.pone.0144649

Happé, F. (1994). Wechsler IQ Profile and Theory of Mind in Autism: A Research Note. *Journal of Child Psychology and Psychiatry, 35*(8) 1461-1471. https://doi.org/10.1111/j.1469-7610.1994.tb01287.x

Happé, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in cognitive sciences, 3*(6), 216-222.

Happé, F. (2018). Why are savant skills and special talents associated with autism? *World Psychiatry, 17*(3), 280. https://doi.org/10.1002/wps.20552

Happé, F., & Frith, U. (1996). The neuropsychology of autism. *Brain, 119*(4), 1377–1400, https://doi.org/10.1093/brain/119.4.1377

Happé, F., & Frith, U. (2006). The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 36*(1). https://doi.org/10.1007/s10803-005-0039-0

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review, 18*(4), 287-304. https://doi.org/10.1007/s11065-008-9076-8

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365-379. https://doi.org/10.1080/0969594970040304

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning, 7*(6), 836-850. https://doi.org/10.1016/j.cptl.2015.08.001

Harris, S., Handleman, J., Belchic, J., & Glasberg, B. (1995). The Vineland Adaptive Behavior Scales for young children with autism. *Special Services in the Schools*, *10*(1), 45-54. https://doi.org/10.1300/J008v10n01_03

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, *78*(9), 1695-1704. https://doi.org/10.1016/j.techfore.2011.04.005

Hastings, R. P., & Brown, T. (2002). Coping strategies and the impact of challenging behaviors on special educators' burnout. *Mental Retardation, 40*(2), 148-156. https://doi.org/10.1016/j.techfore.2011.04.005

Helps, S., Newsom-Davis, I. C., & Callias, M. (1999). Autism: The teacher's view. *Autism, 3*(3), 287-298. https://doi.org/10.1177/1362361399003003006

Hermelin, B., & O'Connor, N. (1967). Remembering of words by psychotic and subnormal children. *British Journal of Psychology, 58*(3), 213–218. https://doi.org/10.1111/j.2044-8295.1967.tb01075.x

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (editors). (2019). *Cochrane Handbook for Systematic Reviews of Interventions,* Version 6.0 (updated July 2019). Cochrane. www.training.cochrane.org/handbook.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review. 24*(2), 189-233. https://doi.org/10.1016/j.dr.2004.01.001

Hoffman, L., Marquis, J., Poston, D., Summers, J. A., & Turnbull, A. (2006). Assessing family outcomes: psychometric evaluation of the Beach Center Family Quality of Life Scale. *Journal of marriage and family*, *68*(4), 1069-1083.

Howell, M., Bradshaw, J., & Langdon, P. E. (2020a). A Systematic review of behaviour-related outcome assessments for children on the autism spectrum with intellectual disabilities in education settings. *Review Journal of Autism and Developmental Disorders*. Advance online publication. https://doi.org/10.1007/s40489-020-00205-y

Howell, M., Bradshaw, J., & Langdon, P. E. (2020b). Asking the teachers: A Delphi study on the selection of skills and behaviours for an assessment of barriers to learning for pupils on the

autism spectrum with intellectual disabilities. *British Journal of Learning Disabilities*. Advance online publication. https://doi.org/10.1111/bld.12350

Howell, M., Bradshaw, J., & Langdon, P. E. (in press). 'There isn't a checklist in the world that's got that on it': Special needs teachers' opinions on the assessment and teaching priorities of pupils on the autism spectrum. *Journal of Intellectual Disabilities.* Advance online publication.

Howlin, P. (2005). Outcomes in autism spectrum disorders. In F. R. Volkmar, A. Klin, R. Paul, & D. J. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (pp. 201–222). Wiley.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2009). Savant skills in autism: psychometric approaches and parental reports. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1522), 1359-1367. https://doi.org/10.1098/rstb.2008.0328

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8. https://doi.org/10.7275/pdz9-th90

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233-253. https://doi.org/10.1111/j.2044-835X.1998.tb00921.x

Hughes, G. (2014). *Ipsative Assessment: Motivation Through Marking Progress*. Palgrave Macmillan.

Hurley, A.D., & Levitas, A.S. (2007). The importance of recognizing autism spectrum disorders in intellectual disability. *Mental Health Aspects of Developmental Disabilities, 10*(4), 57-61.

Hwang, J. L., Davies, P. L., Taylor, M. P., & Gavin, W. J. (2002). Validation of School Function Assessment with elementary school children. *OTJR: Occupation, Participation and Health, 22*(2), 48-58. https://doi.org/10.1177/153944920202200202

Idring, S., Rai, D., Dal, H., Dalman, C., Sturm, H., Zander, E., & Magnusson, C. (2012). Autism spectrum disorders in the Stockholm Youth Cohort: design, prevalence and validity. *PloS one*, *7*(7), 1-9. https://doi.org/10.1371/journal.pone.0041280

Ivanoff, S. D., & Hultberg, J. (2006). Understanding the multiple realities of everyday life: Basic assumptions in focus-group methodology. *Scandinavian Journal of Occupational Therapy*, *13*(2), 125-132. https://doi.org/10.1080/11038120600691082

Janniro, F., Sapp, G. L., & Kohler, M. P. (1994). Validating the Street Survival Skills Questionnaire. *Psychological Reports*, *74*(1), 191-194. https://doi.org/10.2466/pr0.1994.74.1.191

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199-218. https://doi.org/10.1086/376806

Jasper, M. A. (2005). Using reflective writing within research. *Journal of Research in Nursing*, *10*(3), 247-260. https://doi.org/10.1177/174498710501000303

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement*, *70*(3), 394-400. https://doi.org/10.1177/0013164409355692

Johnson, S., & Johnson, R. (2009). *Conceptualising and Interpreting Reliability.* The Office of Qualifications and Examinations Regulation.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26. https://doi.org/10.3102/0013189X033007014

Joosten, A. V., & Bundy, A. C. (2010). Sensory processing and stereotypical and repetitive behaviour in children with autism and intellectual disability. *Australian Occupational Therapy Journal*, *57*(6), 366-372. https://doi.org/10.1111/j.1440-1630.2009.00835.x

Jordan, R. (2001). *Autism with Severe Learning Disabilities.* Souvenir Press (Educational and Academic) Ltd.

Journal of Applied Research in Intellectual Disabilities. (2019, December). *Journal of Applied Research in Intellectual Disabilities: JARID Author Guidelines.* https://onlinelibrary.wiley.com/page/journal/14683148/homepage/forauthors.html

Kaat, A. J., Lecavalier, L., & Aman, M. G. (2014). Validity of the aberrant behavior checklist in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*(5), 1103-1116. https://doi.org/10.1007/s10803-013-1970-0

Kaehne, A., & O'Connell, C. (2010). Focus groups with people with learning disabilities. *Journal of Intellectual Disabilities, 14*(2), 133-145. https://doi.org/10.1177/1744629510381939

Kanne, S. M., Gerber, A. J., Quirmbach, L. M., Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2011). The role of adaptive behavior in autism spectrum disorders: Implications for functional outcome. *Journal of Autism and Developmental Disorders, 41*(8), 1007-1018. https://doi.org/10.1007/s10803-010-1126-4

Kanne, S. M., Mazurek, M. O., Sikora, D., Bellando, J., Branum-Martin, L., Handen, B., & Warren, Z. (2014). The Autism Impact Measure (AIM): initial development of a new tool for treatment outcome measurement. *Journal of Autism and Developmental Disorders, 44*(1), 168-179. https://doi.org/10.1007/s10803-013-1862-3

Kanner, L. (1943). Autistic Disturbances of Affective Contact. *Nervous Child, 2*(3), 217-250.

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

Kern, J. K., Garver, C. R., Grannemann, B. D., Trivedi, M. H., Carmody, T., Andrews, A. A., & Mehta, J. A. (2007). Response to vestibular sensory events in autism. *Research in Autism Spectrum Disorders, 1*(1), 67-74. https://doi.org/10.1016/j.rasd.2006.07.006

Kicklighter, R. H., & Bailey, B. S. (1980). A direct measure of adaptive behavior. *School Psychology Review, 9*(2), 168-73. https://doi.org/10.1080/02796015.1980.12086545

King, N. (2012). "Doing Template Analysis" in Symon, G. & Cassell, C. (Eds.). *Qualitative organizational research: core methods and current challenges*. 426-450. Sage.

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *BMJ, 311*(7000), 299-302. https://doi.org/10.1136/bmj.311.7000.299

Koegel, L., Matos-Freden, R., Lang, R., & Koegel, R. (2012). Interventions for children with autism spectrum disorders in inclusive school settings. *Cognitive and Behavioral Practice, 19*(3), 401-412. https://doi.org/10.1016/j.cbpra.2010.11.003

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kramer, J. M., Coster, W., Kao, Y., Snow, A. & Orsmond, G. (2012). A new approach to the measurement of adaptive behavior: development of the PEDI-CAT for children and youth with autism spectrum disorders. *Physical and Occupational Therapy in Pediatrics*, *32*(1), 34-47. https://doi.org/10.3109/01942638.2011.606260

Kramer, J. M., & Schwartz, A. E. (2017). Refining the Pediatric Evaluation of Disability Inventory–Patient-Reported Outcome (PEDI-PRO) item candidates: interpretation of a self-reported outcome measure of functional performance by young people with neurodevelopmental disabilities. *Developmental Medicine and Child Neurology*, *59*(10), 1083-1088. https://doi.org/10.1111/dmcn.13482

Kramer, J. M., & Schwartz, A. E. (2018). Development of the Pediatric Disability Inventory-Patient Reported Outcome (PEDI-PRO) measurement conceptual framework and item candidates. *Scandinavian Journal of Occupational Therapy*, 1-12. https://doi.org/10.1080/11038128.2018.1502344

Kwok, E. Y., Brown, H. M., Smyth, R. E., & Cardy, J. O. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202-222. https://doi.org/10.1016/j.rasd.2014.10.008

Lam, K. S., & Aman, M. G. (2007). The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(5), 855-866. https://doi.org/10.1007/s10803-006-0213-z

Lane, B. R., Paynter, J., & Sharman, R. (2013). Parent and teacher ratings of adaptive and challenging behaviours in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*(10), 1196-1203. https://doi.org/10.1016/j.rasd.2013.07.011

Lane, A. E., Young, R. L., Baker, A. E., & Angley, M. T. (2010). Sensory processing subtypes in autism: Association with adaptive behavior. *Journal of Autism and Developmental Disorders*, *40*(1), 112-122. https://doi.org/10.1007/s10803-009-0840-2

Law, H., & Morrison, A. P. (2014). Recovery in psychosis: a Delphi study with experts by experience. *Schizophrenia Bulletin*, *40*(6), 1347-1355. https://doi.org/10.1093/schbul/sbu047

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. *Psychological Bulletin, 137*(4), 562-593. https://doi.org/10.1037/a0023341

Leppink, J., & Pérez-Fuster, P. (2017). We need more replication research–A case for test-retest reliability. *Perspectives on Medical Education, 6*(3), 158-164. https://doi.org/10.1007/s40037-017-0347-z

Lewis, A., Lindsay, G., & Phillips, E. (2003). Assessment in special schools: national early assessment procedures and pupils attending special schools in England. *European Journal of Special Needs Education, 18*(2), 141-153. https://doi.org/10.1080/0885625032000078943

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders, 35*(4), 445-460. https://doi.org/10.1007/s10803-005-5035-x

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education, 6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., & Taylor, J. L. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers, 6*(1), 1-23. https://doi.org/10.1038/s41572-019-0138-4

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. https://doi.org/10.1023/A:1005592401947

Ludlow, A., Skelly, C., & Rohleder, P. (2012). Challenges faced by parents of children diagnosed with autism spectrum disorder. *Journal of Health Psychology, 17*(5), 702-711. https://doi.org/10.1177/1359105311422955

Magiati, I., Moss, J., Yates, R., Charman, T., & Howlin, P. (2011). Is the Autism Treatment Evaluation Checklist a useful tool for monitoring progress in children with autism spectrum disorders? *Journal of Intellectual Disability Research*, *55*(3), 302-312. https://doi.org/10.1111/j.1365-2788.2010.01359.x

Male, D. B. (1996). Who goes to SLD schools? *Journal of Applied Research in Intellectual Disabilities*, *9*(4), 307-323. https://doi.org/10.1111/j.1468-3148.1996.tb00117.x

Male, D. (2003). Challenging behaviour: the perceptions of teachers of children and young people with severe learning disabilities. *Journal of Research in Special Educational Needs*, *3*(3), 162-171. https://doi.org/10.1111/1471-3802.00011

Mandell, D. S., Maytali, M., Novak, M. A., & Zubritsky, C. D. (2005). Factors associated with age of diagnosis among children with Autism Spectrum Disorders. *Pediatrics*, *116*(6), 1480–1486. https://doi.org/10.1542/peds.2005-0185

Marshall-Tate, K. (2019). Autism Legislation and Policy. In Chaplin, E., McCarthy, J., & Spain, D. (Eds.) *A Clinician's Guide to Mental Health Conditions in Adults with Autism Spectrum Disorders: Assessment and Interventions* (pp.40-51). Jessica Kingsley Publishers.

Marshburn, E. C., & Aman, M. G. (1992). Factor validity and norms for the Aberrant Behavior Checklist in a community sample of children with mental retardation. *Journal of Autism and Developmental Disorders*, *22*(3), 357-373. https://doi.org/10.1007/bf01048240

Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. *Neuroscience Bulletin*, *33*(2), 183–193. https://doi.org/10.1007/s12264-017-0100-y

Matson, J. L., Belva, B., Hattier, M. A., & Matson, M. L. (2011). Pica in persons with developmental disabilities: Characteristics, diagnosis, and assessment. *Research in Autism Spectrum Disorders*, *5*(4), 1459-1464. https://doi.org/10.1016/j.rasd.2011.02.006

Matson, J., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30*, 107–1114. https://doi.org/10.1016/j.ridd.2009.06.003

Matsuzaki, J., Kuschner, E. S., Blaskey, L., Bloy, L., Kim, M., Ku, M., Edgar, J. C., Embick, D., & Roberts, T. P. (2019). Abnormal auditory mismatch fields are associated with communication impairment in both verbal and minimally verbal/nonverbal children who have autism spectrum disorder. *Autism Research, 12*(8), 1225-1235. https://doi.org/10.1002/aur.2136

Mayfield, K. L., Forman, S. G., & Nagle, R. J. (1984). Reliability of the AAMD Adaptive Behavior Scale - Public School Version. *Journal of School Psychology*, *22*(1), 53-61. https://doi.org/10.1016/0022-4405(84)90051-7

McAllister, K., & Maguire, B. (2012). Design considerations for the autism spectrum disorder-friendly Key Stage 1 classroom. *Support for Learning*, *27*(3), 103-112. https://doi.org/10.1111/j.1467-9604.2012.01525.x

McAllister, K., & Sloan, S. (2016). Designed by the pupils, for the pupils: an autism-friendly school. *British Journal of Special Education*, *43*(4), 330-357. https://doi.org/10.1111/1467-8578.12160

McClintock, K., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: a meta-analytic study. *Journal of Intellectual Disability Research*, *47*(6), 405-416. https://doi.org/10.1046/j.1365-2788.2003.00517.x

McConachie, H., Livingstone, N., Morris, C., Beresford, B., Le Couteur, A., Gringras, P., Garland, D., Jones, G., Macdonald, G., Williams, K., & Parr, J. R. (2018). Parents suggest which indicators of progress and outcomes should be measured in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(4), 1041-1051. https://doi.org/10.1007/s10803-017-3282-2

McConachie, H., Parr, J. R., Glod, M., Hanratty, J., Livingstone, N., Oono, I. P., Robalino, S., Baird, G., Beresford, B., Charman, T., Garland, D., Green, J., Gringras, P., Jones, G., Law, J., Le Couteur, A. S., Macdonald, G., McColl, E. M., Morris, C., Rodgers, J., Simonoff, E., Terwee, C. B. & Williams, K. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technology Assessment*, *19*(41), 1-506. https://doi.org/10.3310/hta19410

McIntosh, J. (2015). *Final report of the Commission on Assessment without Levels, September 2015.* Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf [Accessed 10 October 2017]

McNicholas, J. (2000) Assessment: The Assessment of Pupils with Profound and Multiple Learning Difficulties. *British Journal of Special Education, 27*(3), 150-153. https://doi.org/10.1111/1467-8527.00178

Meilleur A.-A. S., Jelenic P., & Mottron L. (2015). Prevalence of clinically and empirically defined talents and strengths in Autism. *Journal of Autism and Developmental Disorders, 45,* 1354-1367. https://doi.org/10.1007/s10803-014-2296-2

Merrell, K. W. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. *Exceptionality, 9*(1-2), 3-18. https://doi.org/10.1080/09362835.2001.9666988

Mills, C., & Chapparo, C. (2018). Listening to teachers: Views on delivery of a classroom based sensory intervention for students with autism. *Australian Occupational Therapy Journal*, *65*(1), 15-24. https://doi.org/10.1111/1440-1630.12381

Milton, D. (2012). *"So what exactly is autism?"* Autism Education Trust. http://www.aettraininghubs.org.uk/wp-content/uploads/2012/08/1_So-what-exactly-is-autism.pdf [accessed 31/05/18]

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*, *6*(7), 1-6. https://doi.org/10.1371/journal.pmed.1000097

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Brazilian Journal of Physical Therapy, *20*(2), 105-113. https://doi.org/10.1590/bjpt-rbf.2014.0143

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of Bias checklist for systematic reviews of patient-

reported outcome measures. *Quality of Life Research, 27*(5), 1171-1179. https://doi.org/10.1007/s11136-017-1765-4

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2018). Scale development: ten main limitations and recommendations to improve future research practices. Psicologia: Reflexão e Crítica, *30*(3), 1-20. https://doi.org/10.1186/s41155-016-0057-1

Morgan, D. L. (1998). *The Focus Group Guidebook (Vol. 1).* Sage publications.

Morse, J. M. (Ed.). (1994). *Critical Issues in Qualitative Research Methods*. Sage.

Munir, S. Z., Zaman, S., & McConachie, H. (1999). Development of an independent behaviour assessment scale for Bangladesh. *Journal of Applied Research in Intellectual Disabilities, 12*(3), 241-252. https://doi.org/10.1111/j.1468-3148.1999.tb00080.x

Murray, J. (2013). Likert data: what to use, parametric or non-parametric? *International Journal of Business and Social Science*, *4*(11).

Murphy, G. H., Beadle-Brown, J., Wing, L., Gould, J., Shah, A., & Holmes, N. (2005). Chronicity of challenging behaviours in people with severe intellectual disabilities and/or autism: A total population sample. *Journal of Autism and Developmental Disorders, 35*(4), 405-418. https://doi.org/10.1007/s10803-005-5030-2

Murray, D., Lesser, M., & Lawson, W. (2005). Attention, monotropism and the diagnostic criteria for autism. *Autism, 9*(2), 139-156. https://doi.org/10.1177/1362361305051398

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, *142*(2), 71-89. https://doi.org/10.1080/00221309.2014.994590

National Autistic Society. (2013). *Push for Action.* National Autistic Society.

National Institute for Health and Care Excellence. (2018). "*Learning disabilities and behaviour that challenges: service design and delivery*." (NG93). www.nice.org.uk/guidance/ng93 [accessed 22 September 2020].

Neil, L., Green, D., & Pellicano, E. (2017). The psychometric properties of a new measure of sensory behaviors in autistic children. *Journal of Autism and Developmental Disorders*, *47*(4), 1261-1268. https://doi.org/10.1007/s10803-016-3018-8

Nelson, P. B., & Soli, S. (2000). Acoustical barriers to learning: Children at risk in every classroom. *Language, Speech, and Hearing Services in Schools*, *31*(4), 356-361. https://doi.org/10.1044/0161-1461.3104.356

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149-155. https://doi.org/10.1016/j.jrp.2016.06.020

Nicholls, G., Hastings, R. P., & Grindle, C. (2020). Prevalence and correlates of challenging behaviour in children and young people in a special school setting. *European Journal of Special Needs Education*, *35*(1), 40-54. https://doi.org/10.1080/08856257.2019.1607659

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing*, *18*(2), 34-35. http://doi.org/10.1136/eb-2015-102054

O'Brien, G., & Pearson, J. (2004). Autism and Learning Disability. *Autism. 8*(2), 125-140. https://doi.org/10.1177/1362361304042718

Office for Standards in Education, Child's Services and Skills. (2020). *School Inspection Handbook*. Ofsted.

Oliver, C., McClintock, K., Hall, S., Smith, M., Dagnan, D., & Stenfert-Kroese, B. (2003). Assessing the severity of challenging behaviour: psychometric properties of the challenging behaviour interview. *Journal of Applied Research in Intellectual Disabilities*, *16*(1), 53-61. https://doi.org/10.1046/j.1468-3148.2003.00145.x

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, *8*(5), 375-387. https://doi.org/10.1080/13645570500402447

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, *13*(4), 695-705.

Ozonoff, S., Pennington, B.F., & Rogers, S.J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry. 32*(7), 1081-1105. https://doi.org/10.1111/j.1469-7610.1991.tb00351.x

Pan, X., Totsika, V., Nicholls, G., & Paris, A. (2019). Validating GO 4 KIDDS as a brief measure of adaptive skills in special education settings for children with severe intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, *32*(2), 280-287. https://doi.org/10.1111/jar.12524

Pan, S. Q., Vega, M., Vella, A. J., Archer, B. H., & Parlett, G. R. (1996). A mini-Delphi approach: An improvement on single round techniques. *Progress in Tourism and Hospitality Research*, *2*(1), 27-39. https://doi.org/10.1002/(SICI)1099-1603(199603)2:1<27::AID-PTH29>3.0.CO;2-P

Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, *20*(2), 65-77. https://doi.org/10.1016/j.pse.2014.11.002

Pellegrino, J. W. (2017). The two disciplines problem – 'it's like Déjà vu all over again!'. *Assessment in Education: Principles, Policy & Practice*, *24*(3), 359-368. https://doi.org/10.1080/0969594X.2017.1326876

Pellicano, E., Dinsmore, A., & Charman, T. (2014). What should autism research focus upon? Community views and priorities from the United Kingdom. *Autism*, *18*(7), 756-770. https://doi.org/10.1177/1362361314529627

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "Weak" central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology*, *18*(1), 77-98. http://doi.org/10.1017/S0954579406060056

Perry, A., & Factor, D. C. (1989). Psychometric validity and clinical usefulness of the Vineland Adaptive Behavior Scales and the AAMD Adaptive Behavior Scale for an autistic sample. *Journal of Autism and Developmental Disorders*, *19*(1), 41-55. https://doi.org/10.1007/BF02212717

Perry, A., Taheri, A., Ting, V., & Weiss, J. (2015). The GO 4 KIDDS Brief Adaptive Scale. *Journal of Applied Research in Intellectual Disabilities*, *28*(6), 594-597. https://doi.org/10.1111/jar.12143

Persicke, A., Bishop, M. R., Coffman, C. M., Najdowski, A. C., Tarbox, J., Chi, K., & Ranick, J. (2014). Evaluation of the concurrent validity of a skills assessment for autism treatment. *Research in Autism Spectrum Disorders, 8*(3), 281-285. https://doi.org/10.1016/j.rasd.2013.12.011

Pickard, H., Rijsdijk, F., Happé, F., & Mandy, W. (2017). Are social and communication difficulties a risk factor for the development of social anxiety? *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(4), 344-351. https://doi.org/10.1016/j.jaac.2017.01.007

Poet, H., Sharp, C., Garry, J., Harland, J., Keightley, G., Kirkup, C., & Martin, K. (2018). *Assessment Without Levels: Qualitative Research.* Department for Education. https://dera.ioe.ac.uk/32634/1/NFER_AWL_report.pdf [Accessed 24 January 2020]

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Provenzani, U., Fusar-Poli, L., Brondino, N., Damiani, S., Vercesi, M., Meyer, N., & Politi, P. (2020). What are we targeting when we treat autism spectrum disorder? A systematic review of 406 clinical trials. *Autism*, *24*(2), 274. https://doi.org/10.1177/1362361319854641

Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., & Altman, D. G. (2017). Identifying approaches for assessing methodological and reporting quality of systematic reviews: a descriptive study. *Systematic Reviews, 6*(1), 1-12. https://doi.org/10.1186/s13643-017-0507-6

Queen, J. P., Quinn, G. P., & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press.

Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the Nutrition Society*, *63*(4), 655-660. https://doi.org/10.1079/PNS2004399

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. Routledge.

Reed, P., Osborne, L. A., & Waddington, E. M. (2012). A comparative study of the impact of mainstream and special school placement on the behaviour of children with autism spectrum disorders. *British Educational Research Journal*, *38*(5), 749-763. https://doi.org/10.1080/01411926.2011.580048

Resnick, M. (2017). *Lifelong Kindergarten: Cultivating Creativity through Projects, Passion, Peers and Play.* MIT Press.

Rist, R. C. (1977). On the relations among educational research paradigms: From disdain to détente. *Anthropology & Education Quarterly*, *8*(2), 42-49. https://doi.org/10.1525/aeq.1977.8.2.05x1394p

Rivard, M., Theroux, A., Mercier, C., & Parent-Boursier, C. (2015). Indicators of intellectual disabilities in young children with autism spectrum disorders. *Journal of Autism & Developmental Disorders*, *45*(1), 127-137. https://doi.org/10.1007/s10803-014-2198-3

Rochford, D. (2016) *The Rochford Review: final report. Review of assessment for pupils working below the standard of national curriculum tests.* Standards and Testing Agency. https://www.gov.uk/government/publications/rochford-review-final-report [Accessed 29 September 2017]

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

Rouse, M., & Agbenu, R. (1998). Assessment and special educational needs: teachers' dilemmas. *British Journal of Special Education, 25*(2), 81-87. https://doi.org/10.1111/1467-8527.t01-1-00062

Royal College of Psychiatrists, British Psychological Society, Royal College of Speech and Language Therapists. (2007). *Challenging Behaviour: A Unified Approach* (CR144). Royal College of Psychiatrists. http://www.rcpsych.ac.uk/ files/pdfversion/cr144.pdf [Accessed 29 November 2020]

Ruble, L. A., & Dalrymple, N. J. (1996). An alternative view of outcome in autism. *Focus on Autism and Other Developmental Disabilities*, *11*(1), 3-14. https://doi.org/10.1177/108835769601100102

Rumsey, J. M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders*, *15*(1), 23-36. https://doi.org/10.1007/BF01837896

Rumsey, J. M., & Hamburger, S.D. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. *Journal of Clinical and Experimental Neuropsychology*, *10*(2), 201-221. https://doi.org/10.1080/01688638808408236

Schaaf, R. C., Toth-Cohen, S., Johnson, S. L., Outten, G., & Benevides, T. W. (2011). The everyday routines of families of children with autism: Examining the impact of sensory processing difficulties on the family. *Autism*, *15*(3), 373-389. https://doi.org/10.1177/1362361310386505

Schalock, R. L. (2004). The concept of quality of life: what we know and do not know. *Journal of Intellectual Disability Research*, *48*(3), 203-216. https://doi.org/10.1111/j.1365-2788.2003.00558.x

Schalock, R. L. (2011). The evolving understanding of the construct of intellectual disability. *Journal of Intellectual and Developmental Disability*, *36*(4), 227-237, https://doi.org/10.3109/13668250.2011.624087

Schalock, R. L., Luckasson, R., Tassé, M. J., & Verdugo, M. A. (2018). A holistic theoretical approach to intellectual disability: Going beyond the four current perspectives. *Intellectual and developmental disabilities*, *56*(2), 79-89. https://doi.org/10.1352/1934-9556-56.2.79

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2019). The contemporary view of intellectual and developmental disabilities: Implications for psychologists. *Psicothema*, *31*(3), 223-228.

Semino, S., Ring, M., Bowler, D. M., & Gaigg, S. B. (2018). The Influence of task Demands, Verbal Ability and Executive Functions on Item and Source Memory in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(1), 184–197. https://doi.org/10.1007/s10803-017-3299-6

Sethi, C., Harrop, C., Zhang, W., Pritchett, J., Whitten, A., & Boyd, B. A. (2019). Parent and professional perspectives on behavioral inflexibility in autism spectrum disorders: A qualitative study. *Autism, 23*(5), 1236-1248. https://doi.org/10.1177/1362361318810217

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry. 24*(4), 613-620. https://doi.org/10.1111/j.1469-7610.1983.tb00137.x

Siegel, M., Milligan, B., Stein, H., Teer, O., & Smith, K. A. (2013). Telephone administration of the aberrant behavior checklist: A pilot study of feasibility in children with intellectual disability and autism. *Journal of Intellectual Disabilities*, *17*(3), 265-271. https://doi.org/10.1177/1744629513500007

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107. https://doi.org/10.1007/S11336-008-9101-0

Silverman, D. (Ed.). (2016). *Qualitative Research*. Sage.

Silverman, W. P., Silver, E. J., Lubin, R. A., & Sersen, E. A. (1983). Structure of the Minnesota Developmental Programming System Behavioral Scales, Alternate Form C. *American Journal of Mental Deficiency*, *88*(2), 170-176.

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, *6*(1), 1-21.

Smith, A. (2009). The empathy imbalance hypothesis of autism: a theoretical approach to cognitive and emotional empathy in autistic development. *The Psychological Record*, *59*(3), 489-510. https://doi.org/10.1007/BF03395675

Smith, E., Critten, V., & Vardill, R. (2020). Assessing progress in children with severe/profound intellectual disabilities: what are the issues? *Disability & Society*, 1-5. https://doi.org/10.1080/09687599.2020.1719042

Spain, D., Sin, J., Linder, K. B., McMahon, J., & Happé, F. (2018). Social anxiety in autism spectrum disorder: A systematic review. *Research in Autism Spectrum Disorders*, *52*, 51-68. https://doi.org/10.1016/j.rasd.2018.04.007

Sparrow, S., & Cicchetti, D. (1978). Behavior rating inventory for moderately, severely, and profoundly retarded persons. *American Journal of Mental Deficiency*, *82*(4), 365-74.

Spreat, S. (1982). An empirical analysis of item weighting on the adaptive behavior scale. *American Journal of Mental Deficiency*, 87(2), 159-63.

Standards and Testing Agency. (2019). *The Early Years Foundation Stage Profile 2020 Handbook.* Department for Education.

Standards and Testing Agency. (2020a). *The Engagement Model.* Department for Education.

Standards and Testing Agency. (2020b). *Pre-key stage 1: pupils working below the national curriculum assessment standard for the 2020/21 academic year onwards.* Department for Education.

Standards and Testing Agency. (2020c). *Pre-key stage 2: pupils working below the national curriculum assessment standard for the 2020/21 academic year onwards.* Department for Education.

Stedman, A., Taylor, B., Erard, M., Peura, C., & Siegel, M. (2019). Are children severely affected by autism spectrum disorder underrepresented in treatment studies? An analysis of the literature. *Journal of Autism and Developmental Disorders, 49*(4), 1378-1390. https://doi.org/10.1007/s10803-018-3844-y

Steel, M.J.G., Gorman, R. and Flexman, J.E. (1984). Neuropsychiatric testing in an autistic mathematical idiot-savant: Evidence for nonverbal abstract capacity. *Journal of the American Academy of Child Psychiatry, 23*(6), 704-707. https://doi.org/10.1016/S0002-7138(09)60540-9

Stone Fish, L., & Busby, D. M. (2005). The Delphi method. In Sprenkle, D. H., & Piercy, F. P. (Eds.). *Research Methods in Family Therapy* (pp238-253). Guildford Press.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*(3), 217-222. https://doi.org/10.1207/S15327752JPA8003_01

Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing, 70*(9), 1970-1979. https://doi.org/10.1111/jan.12402

Suarez, M. A. (2012). Sensory processing in children with autism spectrum disorders and impact on functioning. *Pediatric Clinics, 59*(1), 203-214. https://doi.org/10.1016/j.pcl.2011.10.012

Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education*, *5*(4), 541-542. http://dx.doi.org/10.4300/JGME-5-4-18

Syriopoulou-Delli, C. K., Cassimos, D. C., & Polychronopoulou, S. A. (2016). Collaboration between teachers and parents of children with ASD on issues of education. *Research in developmental disabilities*, *55*, 330-345. https://doi.org/10.1016/j.ridd.2016.04.011

Szatmari, P., Georgiades, S., Bryson, S., Zwaigenbaum, L., Roberts, W., Mahoney, W. and Tuff, L. (2006). Investigating the structure of the restricted, repetitive behaviours and interests domain of autism. *Journal of Child Psychology and Psychiatry*, *47*(6), 582-590.

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science*, *16*(6), 311-315. https://doi.org/10.1111/j.1467-8721.2007.00527.x

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tay, L., & Jebb, A. (2017). Scale development. In *S Rogelberg: The SAGE Encyclopedia of Industrial and Organizational Psychology*. SAGE Publications Inc.

Tennent, W. (2020). The assessment of reading comprehension in English primary schools: investigating the validity of the Key Stage 2 reading standard assessment test (SAT). *Education, 3-13*, 1-14. https://doi.org/10.1080/03004279.2020.1735472

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, *27*(5), 1159-1170. https://doi.org/10.1007/s11136-018-1829-0

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *7*, Article 769. https://doi.org/10.3389/fpsyg.2016.00769

Tsermentseli, S., Tabares, J. F., & Kouklari, E. C. (2018). The role of every-day executive function in social impairment and adaptive skills in autism spectrum disorder with intellectual

disability. *Research in Autism Spectrum Disorders, 53,* 1-6. https://doi.org/10.1016/j.rasd.2018.05.006

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, *36*(2), 299-316. https://doi.org/10.1080/01411920902919257

Vermeulen, P. (2012) *Autism as Context Blindness.* AAPP Publishing.

Vermeulen, P. (2015). Context blindness in autism spectrum disorder: Not using the forest to see the trees as trees. *Focus on Autism and Other Developmental Disabilities, 30*(3), 182-192. https://doi.org/10.1177/1088357614528799

Vispoel, W. P., Morris, C. A., & Kilinc, M. (2018). Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods*, *23*(1), 1. https://doi.org/10.1037/met0000107

Voelker, S. L., Shore, D. L., Lee, C. H., & Szuszkiewicz, T. A. (2000). Congruence in parent and teacher ratings of adaptive behavior of low-functioning children. *Journal of Developmental and Physical Disabilities*, *12*(4), 367-376. https://doi.org/10.1023/A:1009436230984

Vogel, C. L. (2008). Classroom design for living and learning with autism. *Autism Asperger's Digest*, May-June. http://www.designhub.it/cometa/wp-content/uploads/2016/10/Classroom-Design-for-Living-and-Learning-with-Autism.doc

Wadgave, U., & Khairnar, M. R. (2016). Parametric tests for Likert scale: For and against. *Asian Journal of Psychiatry, 24,* 67-68. https://doi.org/10.1016/j.ajp.2016.08.016

Walker, M., Worth, J., & Van den Brande, J. (2019). *Teacher Workload Survey 2019.* Department for Education.

Wearmouth, J., & Butler, C. (2020). Special educational needs co-ordinators' perceptions of effective provision for including autistic children in primary and middle schools in England. *Education, 3-13, 48*(3), 258-272. https://doi.org/10.1080/03004279.2019.1664401

Wells, K., Condillac, R., Perry, A., & Factor, D. C. (2009). A comparison of three adaptive behaviour measures in relation to cognitive level and severity of autism. *Journal on Developmental Disabilities, 15*(3), 55-63.

Welsh, P., Rodgers, J., & Honey, E. (2019). Teachers' perceptions of Restricted and Repetitive Behaviours (RRBs) in children with ASD: Attributions, confidence and emotional response. *Research in Developmental Disabilities, 89,* 29-40. https://doi.org/10.1016/j.ridd.2019.01.009

Westby, C. E., Stevens Dominguez, M., & Oetter, P. (1996). A performance/competence model of observational assessment. *Language, Speech, and Hearing Services in Schools, 27*(2), 144-156. https://doi.org/10.1044/0161-1461.2702.144

Wigelsworth, M., Oldfield, J., & Humphrey, N. (2015). Validation of the Wider Outcomes Survey for Teachers (WOST): a measure for assessing the behaviour, relationships and exposure to bullying of children and young people with special educational needs and disabilities (SEND). *Journal of Research in Special Educational Needs, 15*(1), 3-11. https://doi.org/10.1111/1471-3802.12030

Wiliam, D. (2017). Learning and assessment: a long and winding road? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 309-316. https://doi.org/10.1080/0969594X.2017.1338520

Wilkinson, K., & Twist, L. (2010). *Autism and Educational Assessment: UK Policy and Practice.* National Foundation for Educational Research.

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

Wing, L., Gould, J., & Gillberg, C. (2011). Autism spectrum disorders in the DSM-V: Better or worse than the DSM-IV? *Research in Developmental Disabilities, 32*(2), 768–773. https://doi.org/10.1016/j.ridd.2010.11.003

Wing, L., & Potter, D. (2002).'The Epidemiology of Autistic Spectrum: Is the prevalence rising? *Mental Retardation and Developmental Disabilities*, *8*(3), 151-161. https://doi.org/10.1002/mrdd.10029

Wolf-Schein, E. G. (1998). Considerations in assessment of children with severe disabilities including deaf-blindness and autism. *International Journal of Disability, Development and Education, 45*(1), 35-55. https://doi.org/10.1080/1034912980450104

World Health Organization. (2015). *International Statistical Classification of Diseases and Related Health Problems (ICD) (10th Edition).* https://icd.who.int/browse10/2015/en

World Health Organization. (2020). *International Statistical Classification of Diseases and Related Health Problems (ICD) (11th Edition).* https://icd.who.int/browse11/l-m/en

Wu, M., Tam, H. P., & Jen, T. H. (2016). *Classical Test Theory. In: Educational Measurement for Applied Researchers,* (pp.73-39). Springer. https://doi.org/10.1007/978-981-10-3302-5_5

Yvonne Feilzer, M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, *4*(1), 6-16. https://doi.org/10.1177/1558689809349691

Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, *29*(1), 1-10. https://doi.org/10.1186/s41155-016-0040-x