Highlights

- A new autism-specific, special needs teacher assessment was tested
- Results showed excellent test-retest reliability and internal consistency
- The assessment strongly correlated with the Teacher Autism Progress Scale
- Teacher feedback was extremely positive
- Results suggest the ABLE-Autism is a useful and effective assessment

The Preliminary Validity and Reliability Evaluation of the Assessment of Barriers to Learning in Education – Autism

Melanie Howell[1], Tom Bailey[2], Jill Bradshaw[1] and Peter E Langdon[2 3]

[1]Tizard Centre, University of Kent, Canterbury, Kent, UK

[2]Centre for Educational Development, Appraisal and Research (CEDAR), University of Warwick, Coventry, UK

[3]Coventry and Warwickshire Partnership NHS Trust

Correspondence concerning this article should be addressed to: Peter Langdon, Centre for Educational Development, Appraisal and Research (CEDAR), University of Warwick, Coventry, CV4 8UW, United Kingdom, Telephone: +44 (0)2476523638

Email: Peter.Langdon@warwick.ac.uk

Declarations of interest: none

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards and approval of the Tizard Ethics Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: Written informed consent was obtained from the participating teachers and informed consent was sought from the parents of participating pupils either verbal, written or opt-out consent depending upon the school.

Abstract

Background: Few robust, autism-specific outcome assessments have been developed specifically for use by teachers in special schools. The Assessment of Barriers to Learning in Education – Autism (ABLE-Autism) is a newly developed teacher assessment to identify and show progress in barriers to learning for pupils on the autism spectrum with coexisting intellectual disabilities.

Aims: This study aimed to conduct preliminary validity and reliability evaluations of the ABLE-Autism.

Methods and Procedures: Forty-eight autistic pupils attending special schools were assessed using the ABLE-Autism. Multi-level modelling was used to evaluate test-retest reliability, internal consistency and convergent validity with the Teacher Autism Progress Scale.

Outcomes and Results: Results showed excellent test-retest reliability and internal consistency. A large effect size suggested that the ABLE-Autism is strongly correlated with the Teacher Autism Progress Scale. Teacher feedback was extremely positive and suggested that the ABLE-Autism is easily understood by teachers, relevant to autistic pupils in special schools and adequately covers the skills and behaviours that teachers believe are important to assess for these pupils.

Conclusions and Implications: Although further validation is recommended, the preliminary evaluation of the ABLE-Autism suggests that it is a useful and effective outcome assessment for autistic pupils in special schools.

Keywords: *Assessment, autism, measurement properties, special educational needs.*

Credit Author Statement

Melanie Howell: Conceptualisation, Methodology, Formal analysis, Investigation, Resources, Writing – Original draft, Writing – Review and editing, Visualization;

Peter E. Langdon: Conceptualisation, Methodology, Formal analysis, Resources, Writing – Original draft, Writing – Review and editing, Visualization, Supervision;

Jill Bradshaw: Conceptualisation, Methodology, Formal analysis, Resources, Writing – Original draft, Writing – Review and editing, Visualization, Supervision;

Tom Bailey: Software, Resources, Formal analysis, Writing – Review and editing.

Abstract

Background: Few robust, autism-specific outcome assessments have been developed specifically for use by teachers in special schools. The Assessment of Barriers to Learning in Education – Autism (ABLE-Autism) is a newly developed teacher assessment to identify and show progress in barriers to learning for autistic pupils with coexisting intellectual disabilities.

Aims: This study aimed to conduct preliminary validity and reliability evaluations of the ABLE-Autism.

Methods and Procedures: Forty-eight autistic pupils attending special schools were assessed using the ABLE-Autism. Multi-level modelling was used to evaluate test-retest reliability, internal consistency and convergent validity with the Teacher Autism Progress Scale.

Outcomes and Results: Results showed excellent test-retest reliability and internal consistency. A large effect size suggested that the ABLE-Autism is strongly correlated with the Teacher Autism Progress Scale. Teacher feedback was extremely positive and suggested that the ABLE-Autism is easily understood by teachers, relevant to autistic pupils in special schools and adequately covers the skills and behaviours that teachers believe are important to assess for these pupils.

Conclusions and Implications: Although further validation is recommended, the preliminary evaluation of the ABLE-Autism suggests that it is a useful and effective outcome assessment for autistic pupils in special schools.

Keywords: *Assessment, autism, measurement properties, special educational needs.*

What this paper adds?

This study has resulted in the availability of an autism-specific school assessment which is both robust and useful to teachers. 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 ABLE-Autism fills a gap in the list of assessments available for teachers of autistic pupils with coexisting intellectual disabilities in special schools. It addresses skills and behaviours that teachers wish to assess for these pupils and results from this study have suggested that it is useful to special needs teachers for the relevant purposes. This study provided preliminary support for the reliability and validity of the ABLE-Autism, showing sufficient results over a number of measurement properties. In practice, 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. Appropriately identifying the barriers to learning for autistic pupils 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.

The Preliminary Validity and Reliability Evaluation of the Assessment of Barriers to Learning in Education – Autism

1. Introduction

Autistic children with coexisting intellectual disabilities who are educated in special schools often have a range of educational needs due to difficulty with communication, sensory processing differences, restricted and repetitive behaviours, attention and social interaction, and challenging behaviour (Jordan, 2001; Fontil et al., 2019). The interaction between these difficulties and the environment often means that the specific needs of this group of pupils may act as barriers to accessing and engaging in learning (Jordan, 2005). Teachers work with a variety of specific teaching strategies and interventions to overcome these barriers and support these pupils to learn new skills and make progress in areas that improve current and future quality of life. Studies in the US and UK which asked teachers their opinions on the needs of their pupils have found that areas related to the specific needs of autistic pupils were seen as teaching and intervention priorities, while academic development was of lowest concern (Azad and Mandell, 2016; authors, 2020b).

In order to select skills for support and intervention and to monitor pupil progress, it is important that the assessments used are appropriate. However, assessments which are commonly used in special schools are often not developed specifically for the needs of autistic pupils who have coexisting intellectual disabilities (authors, 2020a). This group is usually assessed using generic assessments developed for pupils with a wide range of diagnoses (Arnold and Reed, 2016). A recent systematic review of behaviour-related outcome assessments which can be used with pupils on the autism spectrum in special schools showed that many assessments accessible to teachers have not been evaluated for validity and reliability with appropriate populations, in relevant settings or with teacher respondents (authors, 2020a). Many assessments used in schools are not developed specifically for the needs of autistic pupils and few assessments included in the review were

developed with input and in collaboration with teachers even though this is important to ensure sufficient content validity (authors, 2020a). A small number of assessments included in the review showed potential for measuring outcomes in barriers to learning for autistic pupils in special schools, for example the Teacher Autism Progress Scale (TAPS; Dang et al., 2017) and Autism Treatment Evaluation Checklist (ATEC; Charman et al., 2004; Magiati et al., 2011). Both the TAPS and the ATEC were developed as autism-specific outcome assessments and can both be used by teachers. However, further evaluation of these assessments would be required to establish their appropriateness for use with autistic pupils specifically in special schools. For example, the TAPS is a short assessment designed to show change weekly and further evaluation of additional measurement properties is required. The ATEC, in contrast, is a longer measure which aims to show change over longer periods of time but requires further evaluation with autistic samples with coexisting intellectual disabilities in special school settings. Further evaluation and feedback from teachers on how useful and effective these assessments are when conducted in special schools would also be valuable.

Following the systematic review of available assessments, the Assessment of Barriers to Learning in Education – Autism (ABLE-Autism) was developed in order to provide a solution to the lack of robust, autism outcome assessments developed through collaboration with teachers for use in special schools. The ABLE-Autism is an autism-specific teacher assessment developed to support teachers to identify and assess change and progress in barriers to learning for autistic pupils in special schools. The ABLE-Autism was developed using a three-stage process. The findings of the systematic literature review outlined in Section 1 above led to teacher focus groups in which special needs teachers defined and identified barriers to learning, important progress for autistic pupils and the useful aspects of

assessment tools (authors, 2020b). SEN teachers considered important progress for pupils on the autism spectrum to be related to barriers to learning for these pupils. Restricted and repetitive behaviours (RRBs), sensory needs, functional communication, recognition and regulation of emotions and learning behaviours were noted as likely to impact upon further academic or pre-academic progress made in schools and, therefore, were considered priorities due to their potential impact upon education (authors, 2020b). A list of assessment items was devised based on these focus group results and a Delphi exercise was then conducted where teachers rated the assessment items based on comprehensibility, relevance and comprehensiveness (authors, 2020c). After two rounds, 70 items were endorsed by teachers as relevant and comprehensible and these were included in the ABLE-Autism (authors, 2020c). The input of special needs teachers provided initial face and content validity for the new outcome measure. The selection of items through the Delphi exercise provided initial face and content validity for the skills and behaviours included in the new assessment.

The ABLE-Autism has 70 items and is divided into five subscales: Learning Behaviour Barriers (LBB; 14 items), Restricted and Repetitive Behaviour Barriers (RRBB; 12 items), Emotion and Behaviour Regulation Barriers (EBRB; 16 items), Sensory Barriers (SB; 10 items) and Functional Communication Barriers (FCB; 18 items). The assessment is completed by a teacher who knows the pupil well and each item 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*). Instructions are provided to give teachers guidance on the meaning of these terms. The ABLE-Autism is not a linear assessment meaning that there is no requirement for particular items to be achieved before any other items and any changes in specific skills and behaviours, both positive progress or potential regression, can be measured. The assessment has optional comment boxes

for each item where teachers can note any additional information which may be helpful to them and also features a ‘what that looks like for this child’ box where teachers can personalise the item depending upon the targets or needs of individual pupils. Scores are provided for each subscale including an identification of which items represent a pupil’s primary and secondary barriers to learning. If all subscales are completed, an overall score is provided. Higher scores are indicative of more barriers to learning. Some sample items from the ABLE-Autism are shown in Table 1.

Table 1.

ABLE-Autism Sample Items.

[INSERT TABLE 1 HERE]

The objective of this study was to conduct an initial evaluation of the ABLE-Autism (authors, 2020c).

This paper describes the pilot testing and preliminary validity and reliability evaluation of the ABLE-Autism. The research questions were as follows:

a) What is the evidence suggesting that the ABLE-Autism is a reliable assessment?
b) What is the evidence suggesting that the ABLE-Autism is a valid assessment?
c) To what extend do special needs teachers consider the ABLE-Autism comprehensible, comprehensive and relevant to their autistic pupils?

To answer these questions, the test-retest reliability and internal consistency of the ABLE-Autism were considered as well as convergent validity with the Teacher Autism Progress Scale (TAPS; Dang et al., 2017). Teacher feedback was obtained on the usefulness of the new measure for the specified purposes.

2. Methods

2.1 Participants

A total of 48 pupils aged between 4 and 15 years (M_{age} = 10.4 years; Mdn_{age} = 11 years) were assessed by a convenience sample of 21 teachers and 1 higher level teaching assistant (HLTA) with responsibility for classroom teaching and assessment from 13 schools in England and Wales (for consistency, the HLTA will be referred to as a teacher for the purposes of this research). All pupils attended a special school and had autism as a primary need of their Education, Health and Care Plan. Male pupils accounted for 77.1% of participants (n=37). The number of pupils that each teacher assessed ranged from 1 to 7.

2.2 Design and Procedure

Using a convenience sample, teachers were asked to complete both the ABLE-Autism and the Teacher Autism Progress Scale during the autumn and/or spring term of the 2019/20 academic year (Time 1). Teachers were then asked to complete the ABLE-Autism on a second occasion within two weeks of the first completion (Time 2). Each teacher had known the pupil since at least September and had been teaching that pupil for at least ten weeks at the time of the first assessment. Teachers were also asked to complete a feedback questionnaire on completion of the data collection. The teachers who participated in this evaluation of the ABLE-Autism had not participated in previous assessment development stages of the research.

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 teachers wished to participate. Information about the research was also circulated to headteacher groups as well as posted on social media. Written informed consent was obtained from the participating teachers and informed consent was sought from the

child's parents either verbal, written or opt-out consent depending upon the school. Class teachers created pseudonyms for the pupils before the pupil information and assessment results were sent to the researchers to ensure that no personally identifiable information was shared. Specific data on race/ethnicity and socioeconomic status of pupils and demographic characteristics on teachers were not recorded in accordance with the principles of data minimisation. A favourable ethical opinion for this project was given by XXXX XXXX XXXX.

2.3 Teacher Autism Progress Scale

In order to consider the validity of the ABLE-Autism, comparison to a similar measure was considered appropriate.

The TAPS is a school assessment developed to measure progress for autistic pupils in behaviour, social abilities and functional skills (Dang et al., 2017). It was 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, intended to be conducted more frequently, is scored in the opposite direction, and its items are broader and less specific than the items included in 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 (eg. 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 score indicates better performance (i.e. fewer difficulties). Previous research has shown a there is a significant association between higher TAPS scores and improved social responsiveness as measured by the Social Responsiveness Scale (SRS) and reduced challenging behaviour as measured by the Aberrant Behaviour Checklist (ABC; 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 type and number of areas assessed, is also autism-specific and can be completed by teachers, as well as having been 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 evaluated. 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.

2.4 Data Analysis

2.4.1 Overview

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 when pilot testing (Furr, 2011). Limiting data analysis of scales to nonparametric approaches may be overly restrictive when studies have shown 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 distributions and small sample sizes (Queen et al., 2002; Murray, 2013; Wadgave and Khairnar, 2016). Parametric methods were therefore considered appropriate for this research. The adjusted total and subscale scores approximated a normal

distribution apart from the Sensory Barriers subscale. In practice, data often have slight departures from normality and, in this instance, the skewness and kurtosis were not considered marked and the use of parametric tests was judged as meaningful and appropriate.

Multilevel modelling was used to account for the nesting of data and the dependence within pupil-teacher dyads within schools. The unit of analysis was pupils (level one) and the nesting variable was teachers (level two). The inclusion of a third level, school, was considered, however, as fewer than half the participating schools had assessments completed

by more than one teacher and as the influence of schools and teachers could be considered interchangeable, the model which accounted for teacher effects was considered appropriate. Statistical analyses were conducted using IBM SPSS Version 26.

2.4.2 Missing Data

In total, nine item responses were missing at time 1 and 10 responses were missing at time 2. 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 were missing at both time 1 and time 2.

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, summed scores on and subscale scores on the ABLE-Autism were adjusted by dividing the summed score by the number of completed items to generate a total score accounting for missing responses.

2.4.3 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 Alvorado, 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 suggest that the items 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 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 (i.e. teachers reporting on a number of pupils each) using a three-level model (Level 1 – item level, Level 2 – pupil level, Level 3 – teacher level) outlined in Nezlek (2017).

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Manual suggests good internal consistency is indicated by an alpha value ≥ .70 (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018).

2.4.4 Test-Retest Reliability

Test-retest reliability is considered important in 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 ABLE-Autism a second time for 35 pupils approximately two weeks (range 6-30 days, median 14 days, mean 14.37 days, SD 5.07) 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 potential progress over a half term (i.e. six to eight weeks), a short period over which to conduct test-retest reliability was considered appropriate. The number of items and the fact that over half the participating teachers were assessing a number of pupils each meant that recall effects were minimised.

Intraclass correlation coefficients (r_i) were initially 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 (Lorah, 2018). This was transformed to a correlation coefficient, r_{es}, for easier interpretation. F^2 and r_{es} were calculated and reported for both subscale scores and 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.

2.4.5 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 and f^2 and r_{es} were calculated and reported for the adjusted total (summed) score correlated with the TAPS total score (Lorah, 2018). It was hypothesised that the correlation between the adjusted total score of the ABLE-Autism and the TAPS score was likely to be medium to high.

2.4.6 Teacher Feedback

At the conclusion of the data collection, teachers were asked to complete a short feedback questionnaire. Teachers could complete the questionnaire anonymously online via Qualtrics or return the questionnaire by email. The questionnaire contained four questions asking teachers about the comprehensibility, relevance, comprehensiveness and usefulness of the ABLE-Autism (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). 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 give insight into the responses and to ensure that reasons for scores could be determined. A fifth question also asked teachers if they had any further comments or information they would like to share, with space to comment on features of the assessment that they liked or found useful as well as aspects they think could be changed or improved. A total of 16 teachers responded to the feedback questionnaire giving a response rate of 73%.

3. Results

3.1 Descriptive Statistics

Descriptive statistics of the adjusted Time one scores including mean, median, standard deviation, minimum, and maximum are shown in Table 2 below.

Table 2.

Descriptive Statistics of Adjusted Total and Subscale Scores at Time One.

[INSERT TABLE 2 HERE]

3.2 Internal Consistency

Cronbach's $\boldsymbol{\alpha}$ was calculated for each subscale and the total scale score using the three-level model described in Section 2. For all subscales and for the total score, $\boldsymbol{\alpha} \geq .70$. Values are reported in Table 3.

Table 3.

$\boldsymbol{\alpha}$ *= for Subscales and Total Score Using a 3-Level Model*

[INSERT TABLE 3 HERE]

3.3 Test-Retest Reliability

Based on the formulas described in Section 2, f^2 and r_{es} were calculated for each adjusted subscale score and the adjusted total score. F^2 and r_{es} could not be calculated for the

LBB subscale due to model nonconvergence. The single measures, two-way mixed (absolute agreement) r_i was instead reported for the LBB subscale along with the 95% confidence interval as suggested by Koo and Li (2016). Note that r_i for this subscale, however, does not take account of the nested data. Values are reported in Table 4.

Table 4.

F^2 and r_{es} for Adjusted Total Score and Adjusted Subscale Scores

[INSERT TABLE 4 HERE]

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 (Koo and Li, 2016). Test-retest reliability was therefore considered excellent for all subscales and the total score.

3.4 Convergent Validity

The formulas described in Section 2 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.

3.5 Teacher Feedback

Teacher feedback was considered to be a key part of the research process. As acknowledged in previous stages of this research (authors, 2020a; 2020b), 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 asked to answer the questions outlined in

Table 5 with the opportunity to supplement their answers with further comments if they wished. Questions were similar to those which were asked during the development process (authors, 2020c) and covered key areas of content validity including relevance, comprehensiveness and comprehensibility (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). Teachers had the option to provide feedback anonymously.

As can be seen in Table 5, 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 8 or 9. This initial feedback was extremely positive and the encouraging responses may reflect the fact that special needs teachers were consulted at every stage of the assessment development process.

Table 5.

Descriptive Statistics of Teacher Feedback Scores.

[INSERT TABLE 5 HERE]

Optional comments were provided by 10 teachers and these included comments which confirmed their responses, comments about the physical use of the assessment and suggestions of additions or changes. When considering the optional teacher comments, 28 individual comments remarked positively on the usefulness of the ABLE-Autism 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. 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 ABLE-Autism

is created. Only two teachers made comments that the assessment did not add to their current assessment practices and these comments may reflect the fact that there was a degree of artificiality in using the assessment for the research purposes which will be considered further in Section 4. All other teachers who left optional comments suggested that the assessment would be useful, either wholly or partly, with all or some pupils.

4. Discussion

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 assessment as a whole 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 earlier development stage, the relevance and usefulness of the information provided by individual items must be considered 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 in Section 2.4.3, 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 all r_{es} values and r_i for the LBB subscale all falling within the excellent range. In answer to the first research question, 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 pupils

at special needs schools are usually in small classes and therefore the teachers often know their pupils particularly well. This may mean that teachers are aware of individual pupils' skills, behaviours and abilities at a given time and are consistent with 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 not 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 (authors, 2020c), 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 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 ABLEAutism asks teachers to draw upon wider knowledge of that pupil. We hypothesised a moderate to strong correlation between the ABLE-Autism adjusted total score and the total scores on the TAPS. In considering the second research question, the high r_{es} value, therefore, provides evidence for the validity of the ABLE-Autism and suggests that it assesses areas which teachers identify as potential difficulties for autistic pupils 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 order to determine the uni- or multidimensionality of the scale and subscales.

4.2 Use of the Assessment and Teacher Feedback

Teacher feedback was extremely positive and suggested that the ABLE-Autism 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 autistic pupils, useful to show pupil progress and comprehensive in what was assessed.

When considering teacher feedback, it is necessary to acknowledge that, although teachers may have agreed to participate in the trial of the ABLE-Autism 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 ABLE-Autism on top of their classroom duties and usual pupil assessments. Teachers choosing to use the ABLE-Autism outside of this study 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 didn't consider it appropriate to do so.

It is clear, as with any assessment, that not all teachers will find the ABLE-Autism useful for all pupils at any given time. Although teachers' feedback was extremely 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 ABLE-Autism is as useful as possible for both the teachers who use it and the pupils who are being assessed.

4.3 Limitations

The preliminary pilot testing of the ABLE-Autism was positive and provided initial evidence of the validity and reliability of the new assessment. 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) suggested a minimum sample size of 30 for preliminary pilot studies 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, teachers and parents were told that only pupils with a diagnosis of autism could participate. The autism diagnosis of participating pupils was not checked by the research team (i.e. through independent administration of diagnostic assessment which would not have been possible due to COVID-19 regulations) and the research team relied upon schools and teachers to only select pupils who had an autism diagnosis. However, as pupils are attending special schools, it is extremely unlikely that any pupils do not have an appropriate diagnosis due to placement in special schools involving pupils' individual Education, Health and Care Plans which detail their additional needs including diagnosis. Similarly, additional demographic information (for example age, gender, ethnicity, years of teaching experience) was not collected for participating teachers. Although these demographics were not directly related 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.

Thirdly, teachers were provided with a blank assessment for the second completion and instructed to complete it blind without reference to their previous responses. However, teachers may have had access to their previous responses when completing the assessment for the second time and it is possible that some teachers referred to previous responses even though they were instructed not to.

An additional limitation is that 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.

Finally, evaluating the responsiveness of the ABLE-Autism was originally an important part of this study in response to the findings of XXXX et al. (2020a). Responsiveness is an important measurement property for outcome and progress assessments as it is important to determine whether change can be appropriately captured. Teachers were asked to assess the pupils using the ABLE-Autism a third time 6-8 weeks after the initial completion in order to evaluate the responsiveness of the measure. However, this final stage of the assessment evaluation study 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 UK. It is therefore recommended that evaluations of the responsiveness of the ABLE-Autism are conducted in the future.

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

5. Conclusions and Recommendations

The preliminary evaluations of the ABLE-Autism are promising. The assessment shows good to excellent validity and reliability and received extremely positive feedback from the teachers who used it.

A number of recommendations follow from the current study. Firstly, further validation of assessments is always welcome to ensure that measures used in schools are robust and effective. Replication of this study is recommended to provide further evidence for test-retest reliability, internal consistency and convergent validity. It is also important for further aspects of reliability and validity to be evaluated including factor analysis to provide evidence for structural validity, inter-rater reliability and evaluation of the responsiveness of the measure and its ability to show change and progress. Standardisation procedures may also be appropriate in future, as would comparing assessment norms in different groups of pupils (e.g. pupils on the autism spectrum with differing levels of disability, primary and secondary pupils). Finally, feedback should continue to be sought from the teachers using the ABLEAutism as well as consulting parents on the areas assessed and the progress that may be shown.

Appropriately identifying barriers to learning for autistic pupils with coexisting intellectual disabilities allows teachers to plan specific, individualised interventions and teaching to support pupils to gain skills which 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 pupils, parents and school. Ultimately, reducing barriers to access and engagement in learning opportunities will result in the attainment of new, functional skills and improvements in future outcomes and quality of life.

A key recommendation from this research is that teachers continue to be involved in all stages of the development of new assessments which are used in schools. The high teacher

feedback scores reflect the fact that the ABLE-Autism was developed with input from teachers at every stage, from construct definition and item generation to item selection and assessment evaluation. It is crucial that assessments such as the ABLE-Autism are useful for teachers and include the skills and behaviours that they believe are important to assess for their pupils. The evaluation of aspects of reliability and validity of the ABLE-Autism in the present study is an important step in the development of the new assessment as it provides evidence that the assessment measures what is purports to measure and does so consistently

References

Arnold, S., & Reed, P. (2016). Reading assessments for students with ASD: A survey of summative reading assessments used in special education schools in the UK. British Journal of Special Education, 43(2), 122–141.

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

Berchtold, A. (2016). Test–retest: agreement or reliability?. *Methodological Innovations*, *9*, 2059799116672875.

Bujang, M. A., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Archives of Orofacial Science*, *12*(1).

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

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.

Fontil, L., Gittens, J., Beaudoin, E., & Sladeczek, I. E. (2019). Barriers to and facilitators of successful early school transitions for children with autism spectrum disorders and other developmental disabilities: A systematic review. *Journal of Autism and Developmental Disorders*, 1-16.

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. SAGE Publications Ltd.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), 486.

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, *7*(6), 836-850.

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and psychological measurement*, *70*(3), 394-400.

Jordan, R. (2001). *Autism with Severe Learning Disabilities.* London: Souvenir Press (Educational and Academic) Ltd.

Jordan, R. (2005). Managing autism and Asperger's syndrome in current educational provision. *Pediatric rehabilitation*, *8*(2), 104-112.

Kim, H.Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 38(1):52‗54. doi:10.5395/rde.2013.38.1.52

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), 155163.

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

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(1), 8.

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

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.

Murray, J. (2013). Likert data: what to use, parametric or nonparametric?. *International Journal of Business and Social Science*, *4*(11).

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149-155.

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.

Qin, S., Nelson, L., McLeod, L., Eremenco, S., & Coons, S. J. (2019). Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Quality of Life Research*, *28*(4), 1029-1033.

Queen, J. P., Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

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.

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

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), 11591170.

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*, 769.

Wadgave, U., & Khairnar, M. R. (2016). Parametric tests for Likert scale: For and against. *Asian journal of psychiatry*, *24*, 67.

Table 1.

ABLE-Autism Sample Items.

Subscale and Item Label	Item
LBB5b.	[Pupil] will independently transition around the school following instructions or routine
LBB7.	[Pupil] is willing to try a new or unfamiliar activity
RRBB1a.	[Pupil] will accept changes to a normal routine with some warning
RRBB5.	[Pupil] will accept the usual/preferred options not being available
EBRB5c.	[Pupil] will show that they want something to stop in an appropriate way
EBRB10c.	[Pupil] is able to independently stop or reduce behaviour which may hurt or harm others
SB1.	[Pupil] will recognise when they need sensory input
SB5.	[Pupil] will accept wearing clothes/shoes appropriate to or necessary for the weather/setting
FCB6a.	[Pupil] will appropriately initiate interaction with an adult when something is wanted/needed
FCB10.	[Pupil] will request help appropriately

Table 2.

Descriptive Statistics of Adjusted Total and Subscale Scores at Time One.

Scale/subscale	*M*	*Mdn*	*SD*	*Min*	*Max*

Learning Behaviour Barriers	1.98	1.97	0.91	.14	3.79
Restricted and Repetitive Behaviour Barriers	1.93	1.88	0.80	.33	3.58
Emotion and Behaviour Regulation Barriers	2.02	1.84	0.89	.25	3.63
Sensory Barriers	2.24	2.50	1.01	.00	3.80
Functional Communication Barriers	2.25	2.31	0.83	.83	3.78
Total Score	2.08	2.06	0.75	.71	3.60

Table .

3 ***α*** *= for Subscales and Total Score Using a 3-Level Model*

Scale/subscale	**α**
Learning Behaviour Barriers	.89
Restricted and Repetitive Behaviour Barriers	.86
Emotion and Behaviour Regulation Barriers	.87
Sensory Barriers	.80
Functional Communication Barriers	.88
Total Score	.95

Table . 4

F^2 and r_{es} for Adjusted Total Score and Subscale Scores

Scale/subscale (adjusted scores)	f^2	r_{es}
Learning Behaviour Barriers[a]	Not calculated due to model nonconvergence r_i = .97 (95% CI [0.94, 0.99], p <0.001)	
Restricted and Repetitive Behaviour Barriers	7.92	0.94
Emotion and Behaviour Regulation Barriers	5.92	0.92
Sensory Barriers	5.94	0.93
Functional Communication Barriers	14.75	0.97
Total Score	13.10	0.96

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

5

Descriptive Statistics of Teacher Feedback Scores.

Table .

Question *(1 Not at all – 10 Extremely)*	*M*	*95% Confidence Interval*	*Mdn*	*SD*
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