School readiness in Children with Special Educational Needs and Disabilities

Hughes, Claire H; Foley, Sarah; White, Naomi; Devine, R.T.

DOI:
10.1111/bjep.12206

School Readiness in Children with Special Educational Needs and Disabilities: Psychometric Findings From a New Screening Tool, the Brief Early Skills and Support Index (BESSI)

Claire Hughes

Sarah Foley

Naomi White

Rory T. Devine

Centre for Family Research, University of Cambridge, UK.

Author Note
Correspondence concerning this article should be addressed to Claire Hughes, Centre for family Research, Free School Lane, Cambridge, CB2 3RQ (email: ch288@cam.ac.uk). This study was funded by the Westminster Foundation and the Foundation Years Trust, Birkenhead, UK.

Abstract

Background: There is an urgent need to accelerate the detection of special educational needs and disabilities (SEND). A recent brief questionnaire designed for teachers and nursery staff, the Brief Early Skills and Support Index (BESSI) shows promising psychometric properties (Reference Masked), but has yet to be evaluated as a tool for detecting children who may have SEND.

Aims: Addressing this gap, the current study aimed to assess whether BESSI scores: (i) show measurement invariance across SEND status; (ii) show unique associations with SEND status; and (iii) are sensitive and specific to SEND status.

Sample: 84 teachers and nursery staff completed BESSI ratings for 2106 British children aged 2.5 to 5.5 years (48.9% male, 20% ethnic minority, 9.3% with a statement of SEND).
Method: We applied multi-level confirmatory factor analyses, regression analyses and ROC analyses to examine each of the study questions, using the BESSI subscales (behavioral adjustment, language and cognition, daily living skills and family support) as dependent variables.

Results: The four BESSI subscales were reliable and showed measurement invariance across SEND status. Over and above effects of age, gender, family income, ethnicity and family size, SEND status predicted substantial unique variance in BESSI scores. ROC analyses showed that in detecting children identified as having SEND, a cut-off score of 8.50 on the BESSI total score produced good levels of sensitivity and specificity; gender-specific analyses indicated a lower cut-off score of 6.50 for girls.

Conclusion: The BESSI appears to be a useful tool in screening children for more detailed assessment of SEND.

School Readiness in Children with Special Educational Needs and Disabilities: Psychometric Findings From a New Screening Tool, the Brief Early Skills and Support Index (BESSI)

While the transition to school often acts as a catalyst for identifying childhood disorders, the onset of problems is often much earlier, creating an urgent need for sensitive measures to accelerate the detection of atypical development (Lumu, Albertyn, & Szabo, 2015). In a landmark epidemiological study (Lavigne et al., 1996), showed that approximately 20% of 3- to 5-year-old children met criteria for emotional and behavioral problems but only a quarter of these children had been referred for treatment (Lavigne et al., 1998). As these findings demonstrate, the preschool prevalence of clinical problems is often as high as that for school-aged children (Briggs-Gowan, Carter, Skuban, & Horwitz, 2001; Egger & Angold, 2006); indeed, preschool problems are often precursors to later difficulties (Skovgaard et al., 2008). Adding urgency to the challenge of early detection is growing evidence for environmental influences on children's early development and behavioural adjustment (Shonkoff, 2003). Indeed, effective interventions are now available for numerous forms of developmental delay (McIntyre, 2008). Delayed identification of problems often impedes learning and school engagement (Lyon et al., 2002), making early intervention especially valuable for children with Special Educational Needs and Disabilities (SEND).

In the UK, around 12% of primary school pupils (and 10% of secondary school pupils) receive SEND support (DofE, 2016). Examples of diagnoses that commonly lead to a SEND statement include: emotional and behavioral difficulties, autism spectrum disorders, dyslexia, communication difficulties and medical difficulties such as epilepsy and cerebral palsy. Despite this diversity of conditions, according to Section 20 of the UK Children and Families Act 2014 eligibility for a statement of SEND is defined by one of two key needs: (a) significant difficulties in learning or (b) a disability that limits children's use of mainstream

school facilities. Currently, all children in the UK with a statement of SEND are entitled to a yearly allowance of up to £6,000 to cover the costs of educational support (e.g., one-to-one sessions with a teaching assistant or speech and language therapist), with additional educational health and care (EHC) plans providing access to a second tier of more extensive support if necessary. This governmental support is not only vitally important for the children and their families, but also cost effective: in the USA for example, pre-kindergarten intervention can result in savings of $30-100K per child (Karoly, Kilburn, & Cannon, 2005). Likewise, in the UK, a large-scale randomized controlled trial of speech and language therapy (averaging at 5.5 hours over a 6-month period) showed clear improvements, equivalent to approximately 0.5 SD compared with no-treatment controls, in children's comprehension, expression and speech (Broomfield & Dodd, 2011).

Unfortunately, there are marked discrepancies in the age at which children from different family backgrounds receive appropriate diagnoses (Mandell, Listerud, Levy, & Pinto-Martin, 2002) and so improving the early detection of childhood problems is also important from a societal perspective in ensuring equality of opportunity. In Pennsylvania for example, diagnosis of Autism Spectrum Disorder (ASD) occurs, on average, almost a year later in 'near-poverty' families than in more affluent families (Mandell, Novak, & Zubritsky, 2005). Likewise, a British population study showed that children with less educated parents were the least likely to receive a prior local identification of ASD (Baird et al., 2006), with a recent survey of teachers conducted by the Department of Education also revealing similar family-related variability in the efficiency of SEND referrals and statements (Johnson, Carroll, & Bradley, 2017). Importantly, it is not just the efficiency of the diagnostic process that appears related to family background, but also whether a diagnosis is made; a review of over 1100 speech and language therapy referrals showed children from severely deprived family backgrounds were less likely to receive a diagnosis than their more affluent peers

(Broomfield & Dodd, 2004). There is therefore converging evidence from multiple studies to suggest that disadvantaged families are especially likely to benefit from active screening programmes.

In a recent review, Szaniecki & Barnes (2016) reported that most available measures of infant mental health were developed for research rather than in educational settings and so may be too expensive to implement widely, or may require specialist training. These authors also noted that the clinical utility of screening tools depends upon: (a) reasonable reliability and validity; (b) suitability for diverse families; (c) good sensitivity (i.e., a reasonably low 'false-negative' rate); and (d) good specificity (i.e., a relatively low 'false-positive' rate). Moreover, developmental progress in any domain (e.g., language, motor skills) shows marked normative variation and delay can be manifest differently at different ages. Further, in the early years it is especially important to untangle effects of learning difficulties from those of environmental adversity, as children are entirely dependent on their caregivers.

Given these many measurement challenges, the relative importance of sensitivity versus specificity deserves consideration. Recent shifts towards inclusive educational policies (Runswick-Cole, 2011) have reduced the dangers associated with over-referral, such that sensitivity is probably the most important feature of a good screening instrument. Current detection rates of developmental disorders are consistently lower than their actual prevalence (Sand et al., 2005), indicating that further work is needed to ensure the early identification of children with developmental disorders. If screening instruments are to be implemented widely they should also be quick to complete. Rydz, Shevell, Majnemer and Oskoui (2005) found that time pressure was the most common reason cited by primary care practitioners for failing to administer a standardized developmental checklist. Repeated assessments accentuate the sensitivity of detecting developmental delay or disorder, but increase the need for brief instruments.

A second solution is to move away from parent reports to draw on the experience and skills of preschool educational staff. Interestingly, all six measures of infant mental health identified in Szaniecki and Barnes' (2016) review relied primarily on caregiver report, even though information from nursery staff or teachers is likely to improve the quantity, quality and relevance of ratings, for several reasons. First, limited reading skills may prevent some caregivers from providing full and valid responses to questionnaires (Davis et al., 1994), but these problems are less likely to affect the ratings provided by educational staff. Second, studies of several different groups (e.g., children with ASD or ADHD) have consistently shown an association between parental depression and concern about child (Karp, Ibañez, Warren, & Stone, 2017). Third, peer problems are often a key reason for referring children for SEND evaluation (McKay & Neal, 2009), such that ratings from nursery staff who have ample opportunities to observe children interacting with their peers may prove very useful.

For all of the above reasons, one would expect that a screening measure completed by nursery staff may substantially increase SEND referral rates. To this end, the current study builds on two studies that demonstrate the developmental suitability and reliability of a new and brief measure of children's early skills and support from families (the 30-item Brief Early Skills and Support Index – BESSI; *Reference masked*). The first study began with separate focus groups conducted with Foundation Years teachers and head-teachers, designed to elicit teachers' perspectives on the difficulties displayed by young children making the transition to school. This initial canvassing of teachers' views was useful and led to the construction of a family support subscale, a domain of concern for teachers that has not been included in prior longer measures of school readiness, such as the Early Development Inventory (Janus & Offord, 2007). This was followed by postal surveys to ensure that a wider range of teachers judged the items to be relevant and clearly worded. In the first study, teachers' ratings for a sample of 1,456 children aged 2.5 to 5.5 on the four BESSI subscales (Language and

Cognition, Daily Living Skills, Behavioral Adjustment and Family Support) were sensitive to effects of: (i) age (fewer problems for older children); (ii) gender (more problems for boys); and (iii) family income (more problems for children eligible for free school meals (FSM)/Pupil Premium). Moreover, reduced family support explained the higher level of problems displayed by children eligible for FSM (*Reference masked*). The second study showed that the BESSI has good predictive utility and is sensitive to improvements across a 6-month period, in that group mean scores for problems on all four subscales showed significant reductions over this relatively short time-period (*Reference masked*). In this longitudinal study, teachers' initial ratings of family support predicted unique variance in change scores for the three child subscales of the BESSI (Behavioural Adjustment, Language/Cognition and Daily Living Skills) even when FSM status was taken into account. Whilst we expected contrasts between children with or without a statement of SEND to be primarily related to these three child subscales, we also included the family support subscale as a useful comparator.

Statistical modelling in the two studies showed that BESSI ratings displayed measurement invariance (i.e., similar factor structure and subscale item loadings) across age groups, gender, ethnic minority status and time. Establishing whether BESSI ratings also show measurement invariance across SEND status was an important preliminary step in the current study. Our first main aim was to examine whether SEND status explained unique variance in BESSI scores, even when effects of age, gender, ethnic minority status and family income were taken into account. Related to this aim, we also compared the percentages of children with or without identified SEND rated as having difficulties on each of the BESSI items. Our second main aim was to assess the BESSI's sensitivity and specificity to SEND. Our third aim was to examine whether cut-off scores differed by gender or family background. Given that problems of behavioural adjustment or developmental delay are

generally more common in boys than in girls, we predicted that the cut-off scores would be higher for boys than for girls. Motivated by reports of income-related contrasts in the age at which children are identified as having ASD, we also predicted that cut-off scores would be higher for children eligible for Pupil Premium.

Method

Participants

The current study combined four different datasets collected to examine the psychometric properties of the BESSI (*References Masked*). In the current study we combined questionnaires collected from nurseries and primary schools in the Northwest, East and Southeast of England. From this database of 2205 questionnaires, we excluded 99 cases for whom teachers did not report on the child's special educational needs status. The remaining 2106 children (49.1% boys) were aged 30 to 70 months, M = 52.21 months, SD = 7.37. Based on data from 2074 children, 79.8% of the sample was White British. The four most common ethnic minority groups were Black African (N = 144), Bangladeshi (N = 135), Pakistani (N = 125) and Eastern European (N = 117). Two smaller ethnic minority groups were Black Caribbean (N = 48) and Indian (N = 37). Of the 1856 children for whom data were provided by the teacher, 25% were eligible for Pupil Premium (i.e., children whose parents were in receipt of income support from the state). Of the total sample, 195 children (9.3%) were reported to have a statement of special educational needs and disability (SEND). The data from these children with SEND have not been included in previous publications. Note that we did not have access to detailed information regarding the specific nature of individual children's SEND statements, or to information about the ages at which children had received a SEND statement. The children with SEND, M_{age} = 53.60, SD = 7.13, were older than the typically developing children, M_{age} = 52.06 months, SD = 7.38, t (2098) = 2.73, p = .006, Cohen's d = 0.21 (note that age could not be calculated for 6 typically developing

children). There was a higher proportion of boys in the group of children with SEND (70.8%) than in the typically developing group (46.9%), χ^2 (1) = 40.29, $p < .001$, OR = 2.74, 95%CI [1.99, 3.78]. Children in the group with SEND were twice as likely to be in receipt of Pupil Premium (42.6%) than typically developing children (23.2%), χ^2 (1) = 32.17, $p < .001$, OR = 2.46, 95%CI [1.79, 3.39]. Children with SEND were also more likely to be ethnic minority backgrounds (37%) than children without SEND (18.4%), χ^2 (1) = 37.22, $p < .001$, OR = 2.60, 95%CI [1.89, 3.56], with a similar distribution of specific ethnic minorities across these two groups. Given these differences between the two samples, we included these covariates in our final structural equation models.

Measures

The Brief Early Skills and Support Index (BESSI - Reference Masked). The BESSI is a 30-item questionnaire designed to measure teachers' perceptions of children's school readiness in three child domains (i.e., Behavioral Adjustment, Language and Cognition, Daily Living Skills) and one Family Support domain. Teachers rate the degree to which they agree with each statement about the child on a four-point scale (from Strongly Agree to Strongly Disagree). As described elsewhere (Reference Masked), the BESSI was originally designed to have five 6-item subscales, but exploratory and confirmatory factor analyses revealed a substantial overlap between two of these putative subscales, social relationships and self-regulation. As a result, the 12 items were pooled together to create an index of Behavioral Adjustment (e.g., Is easily distracted). The remaining 3 subscales each had 6 items and indexed: (i) Language and Cognition (e.g., Understands wh-questions); (ii)Daily Living Skills (e.g., Is fully toilet trained); and (iii) Family Support (e.g., Talks about fun, shared activities at home). Items are scored on a binary scale with high scores indicating the presence of a problem in that particular domain. Each subscale of the BESSI shows excellent 1-month test-retest reliability and strong stability over 6 months (*References*

Masked). Teachers also reported on children's age, gender, ethnicity, eligibility for Pupil Premium and whether or not the children had older siblings.

Procedure

Teaching staff distributed information sheets to caregivers, who were able to withdraw their child from the study if they did not want their child's information to be used. Teachers completed the BESSI questionnaire for all or half of their class (i.e., every other child on the register), or for the children for whom they were the key worker (in nursery settings) and received gift vouchers class for taking part.

Results

Analytic Strategy

We used latent variable modelling in M*plus* Version 7 (Muthén & Muthén, 2012) to examine the factor structure, measurement invariance and predictors of performance on the BESSI. Given the categorical nature of the BESSI data, we used mean- and variance-adjusted weighted least squares estimation (Kline, 2012). We evaluated model fit using Brown's (2015) recommended criteria: root mean square error of approximation (RMSEA) < .08, comparative fit index (CFI) > .90, and Tucker Lewis index (TLI) > .90. There was less than 1% item non-response across 29 items of the BESSI. One item (regarding reading at home) had missing data for 8.7% of the sample. We used a full-information approach (in which missing model parameters and standard errors were estimated using all available data) to analyse the data so that all cases could be included in the analyses (Enders, 2001).

Measurement Models

Table 1 shows the item-level performance data for the whole sample and for the children with and without SEND. Table 2 shows the tetra-choric correlations for the BESSI items. We tested the fit of the previously reported four-factor solution for the BESSI (i.e., Behavioral Adjustment, Language and Cognition, Daily Living Skills, and Family Support)

in our combined sample (Model 1). This model provided a good fit to the data (Table 3). All BESSI items loaded significantly onto their respective latent factors and each latent factor exhibited significant variance (see Table 4 for model results listed under 'Single Level'). The reliability coefficient for each latent factor was acceptable for children with average levels of ability and support (i.e., when $\theta = 0$): Behavioral Adjustment, .90, Language and Cognition, .78, Daily Living Skills, .72, and Family Support, .70. The test was most reliable when measuring those children with higher than average numbers of problems (i.e., when $+0.5SD \leq \theta \leq +1SD$): Behavioral Adjustment, .94 - .95, Language and Cognition, .87 - .88, Daily Living Skills, .84 - .85, and Family Support, .86 - .83.

Measurement Invariance and Group Differences

We used multiple-groups CFA to examine the measurement invariance of the BESSI across children with and without SEND and the magnitude of latent mean differences between the two groups. We tested measurement invariance across the two groups by comparing the fit of a sequence of increasingly constrained models beginning with a test of equal form (or configural invariance), then equal factor loadings (or metric invariance), and then equal indicator thresholds (or scalar invariance) (Brown, 2015). We also tested strict factorial invariance by testing the equality of item residuals across both groups (Brown, 2015). Together these tests were used to determine whether the items measured the same latent construct (equal form), whether each item exhibited similar relationships to the underlying constructs across the two groups (equal factor loadings), and whether any items were biased toward either group (equal thresholds) (Brown, 2015). Changes in model fit and inspection of the modification indices were used to assess for the presence of differential item functioning (DIF). We used the corrected χ^2 difference test (suitable for use with mean- and variance-adjusted weighted least squares estimation) to compare nested models. We adjusted our α to compensate for multiple comparisons ($\alpha = 0.05/18 = 0.0028$).

To examine the assumption of equal form, we tested the four latent factor model of the BESSI across the two groups simultaneously using multiple-group CFA. In this baseline model we allowed the item loadings and thresholds to be freely estimated across each group. In order to identify the model, the item residuals, latent factor means, and latent factor variances were constrained to equality across both groups. The model provided an acceptable fit to the data (Model 2, Table 3) confirming the assumption of equal form. Next, to examine the assumption of equal factor loadings, we constrained the factor loadings for each item to be equal to their corresponding item across the two groups and allowed the factor variances to be freely estimated across the two groups. This model provided an acceptable fit to the data and did not result in a substantial decrease in overall model fit (Model 3, Table 3) supporting the assumption of equal form and equal factor loadings. We then assessed the assumption of equal thresholds by constraining the item thresholds to be equal across the two groups and freely estimating the four latent factors means across each group. While this model provided an adequate fit to the data, the χ^2 difference test indicated a substantial decrease in model fit suggesting that the assumption of equal thresholds was not supported (Model 4, Table 3).

To investigate these results further we re-ran our model but permitted the item residual variances to be freely estimated across both groups. This model provided an adequate fit to the data (Model 5, Table 3) and inspection of the modification indices revealed that the threshold constraints on two items were causing strain on the model. We released the constraints on the thresholds for two items on the Language and Cognition subscale (i.e., 'Speaks clearly', 'Enjoys identifying letters') and this produced a model with acceptable fit (Model 6, Table 3). Inspection of the thresholds indicated that teachers were more likely to report problems on the 'Speaks Clearly' item when rating children with SEND (Latent factor threshold = 0.35) even if these children had similar levels of underlying latent ability in

Language and Cognition as children without SEND (Latent factor threshold = 0.78). In contrast, teachers were more likely to identify problems on the 'Enjoys identifying letters' item when rating children without SEND (Latent factor threshold = 0.60) even if these children had similar levels of latent ability as children with SEND (Latent factor threshold = 0.88). To examine the equality of item residuals across the two groups, we compared the fit of Model 6 with a model in which each of the item residuals were constrained to be equal across the two groups. This partial measurement invariance model provided a good fit to the data and was not significantly different from Model 6 (Table 3, Model 7). This confirmed the presence of so-called strict factorial invariance or equality of indicator residuals (Brown, 2015). To summarize, multiple-groups CFA supported the assumptions of equal form, loadings and indicator residuals and partially supported the assumption of equal thresholds.

Having established partial measurement invariance, we tested population heterogeneity by first examining the equivalence of latent factor variances and then evaluating latent mean differences (Brown, 2015) between the children with and without SEND. Using Model 7 (the partial measurement invariance model) as our baseline, we undertook tests of population heterogeneity. Prior to comparing latent means, it was necessary to test whether the latent factor variances were equal across both groups (Brown, 2015). We assessed the equality of latent factor variances across the two groups for each latent factor by comparing the fit of each nested model with the last (Models 8 – 11). With the exception of Language and Cognition, there were no significant decreases in model fit when the latent factor variances for Behavioral Adjustment, Daily Living Skills, and Family Support were set to be equal across both groups (see Table 3, Models 8 - 11). Together these models indicated that the latent factor variances were mostly equivalent across the two groups. We therefore calculated the latent mean differences from the output of Model 11 in which all factor variances were constrained to equality across both groups. The χ^2 difference

tests and model fit indices for the latent mean difference test models are presented in Table 3 (Models 12 – 15). Children with SEND had higher latent factor means than children without SEND on each of the four latent factors: Behavioral Adjustment, *Unstandardized Est*. = 1.057, 95%CI [0.91, 1.21], $p < .0001$, Language and Cognition, *Unstandardized Est*. = 1.129, 95%CI [0.95, 1.31], $p < .0001$, Daily Living Skills, *Unstandardized Est*. = 1.509, 95%CI [1.33, 1.69], $p < .0001$, and Family Support, *Unstandardized Est*. = 0.884, 95%CI [0.70, 1.07], $p < .0001$.

Given that the data were collected from 84 teaching staff, we used multi-level modelling to partition the variance accounted for by child-level variation (i.e., individual differences) and teacher-level variation. The ICCs for 27 items of the BESSI exceeded .10 (Table 1). We therefore specified a two-level CFA to examine the fit of the four-factor model once between-teacher differences were taken into account (Byrne, 2012). To this end we tested a baseline model with four latent factors at the within (individual) and between (teacher) levels (Model 16). This model provided an acceptable fit to the data but inspection of the model parameters revealed that the residual variance for one between-teacher level item ('Speaks clearly') was close to 0. We therefore fixed the residual for this item at 0 on the between-teacher level and re-ran the model (see Model 17, Table 3 for model fit indices and Table 4 for parameter estimates). Next, to test the equality of factor loadings across levels, we constrained the item loadings to be equal on both levels. This model provided a good fit to the data (see Model 18, Table 3). In sum, our multilevel CFAs revealed that the four-factor solution provided an acceptable fit to the data even when potential between-teacher variation was taken into account.

Unique Relations Between BESSI Ratings and SEND Status

As children with and without SEND differed on a range of background variables, we used multi-level structural equation modelling to examine the unique association between

SEND status and performance on each of the BESSI latent factors. We regressed each of the BESSI latent factors (at the within-level only) onto a binary dummy variable indicating the presence (1) or absence (0) of SEND and several covariates: eligibility for Pupil Premium, minority ethnic status, gender, the presence of older siblings, and age in months. Given earlier findings that two items from the Language and Cognition subscale exhibited DIF, we regressed each of these items onto the SEND dummy variable to account for the item non-invariance (Brown, 2015). In doing so, we sought to examine the unique association between SEND status and BESSI latent factor scores. We chose to use the BESSI latent factor scores as the dependent variable as these scores were continuous and not binary like SEND status. This model provided an acceptable fit to the data (Model 19, Table 3). Table 5 shows the unstandardized and standardized estimates for the regression of each latent factor onto the six predictors. Given the large sample size, we selected a more stringent alpha level ($\alpha = .001$).

The model accounted for a significant degree of variance in each of the four BESSI latent factors: Behavioral Adjustment, $R^2 = 17\%$, Language and Cognition, $R^2 = 41\%$, Daily Living Skills, $R^2 = 33\%$, and Family Support, $R^2 = 28\%$. As reported elsewhere (Reference Masked), boys showed more problems than girls on each of the four latent factors. Older children had fewer problems than younger children in Language and Cognition and Daily Living Skills. Children in receipt of Pupil Premium showed more problems than their peers in Language and Cognition, Daily Living Skills and Family Support. For the Language and Cognition and the Family Support subscales, children from ethnic minority families showed more problems than their White British counterparts.

Crucially, children with SEND displayed significantly more problems on each of the four factors of the BESSI even when individual differences in age, gender, family income (Pupil Premium Status), ethnicity and the presence of older siblings were taken into account. SEND status uniquely accounted for 9% of the variance in Behavioral Adjustment, 10% of

the variance in Language and Cognition, 16% of the variance in Daily Living Skills and 7% of the variance in Family Support.

Next we specified a multi-level structural equation model to examine whether SEND status continued to be correlated with individual differences in Behavioral Adjustment, Language and Cognition, and Daily Living Skills once ratings of Family Support were taken into account. To this end we regressed the three BESSI latent factors (at the within-level only) onto a binary dummy variable indicating the presence (1) or absence (0) of SEND, eligibility for Pupil Premium, minority ethnic status, gender, the presence of older siblings, age in months, and the Family Support latent factor scores. This model provided an acceptable fit to the data (Model 20, Table 3). Table 5 shows the parameter estimates for this model. The final model accounted for a significant proportion of the variance in Behavioral Adjustment, R^2 = .33, Language and Cognition, R^2 = .70, and Daily Living Skills, R^2 =.64. Even when individual differences in Family Support were taken into account, SEND status continued was significantly associated with each of the three BESSI latent factors: Behavioral Adjustment, *Std*. *Est*. = .16, 95%CI [.12, .20], $p < .0001$, Language and Cognition, *Std*. *Est*. = .17, 95%CI [.13, .21], $p < .0001$, and Daily Living Skills, *Std*. *Est*. = .24, 95%CI [.19, .28], $p < .0001$. That is, the correlations between SEND status and each of the three child-focused BESSI latent factors were not accounted for by individual differences in children's age, gender, ethnic minority status, family size, family income or family support.

Sensitivity, Specificity and Cut-Off Scores

We performed a receiver operating characteristic (ROC) curve analysis on the summed raw scores of the BESSI to determine the sensitivity and specificity of BESSI total scores as a tool for classifying children as having SEND. While the CFAs supported a four-factor structure, we sought to simplify our analyses by focusing on a single total score for ease of practical application in the field. Table 6 presents the sensitivity and specificity values

for the BESSI at different cut-off scores for the whole sample, by gender and by age. The area under the curve was significant, AUC = .84, 95%CI [.81, .87], $p < .001$, indicating that BESSI total scores performed better than chance when classifying children with SEND (see Figure 1). Looking at the whole sample, cut-off scores of 8.50 provided a sensitivity of .80 and a specificity of .75. Approximately 30% ($N = 630$) of the total sample scored above 8.50 on the BESSI. The positive predictive value (PPV) for this cut-off (i.e., the probability that children scoring above 8.50 have SEND or 'true positives') was .25 and the negative predictive value (NPV) (i.e., the probability that children scoring below 8.50 do not have SEND or 'true negatives') was .97. That is, in a sample of 100 children, 30 would score over 8.50. Of these 30, 75% would be false positives ($N = 23$) and 25% would be true positives ($N = 7$). Of those scoring under 8.50, 97% would be true negatives ($N = 68$) and 3% would be false negatives ($N = 2$).

Inspection of the sensitivity and specificity statistics by age (Table 6) revealed that a cut-off score of 8.50 yielded similar patterns of sensitivity and specificity in each age band. In contrast, analysis of the data by gender indicated that there may be different cut-off scores for girls, AUC = .84, 95%CI [.78, .89], $p < .0001$, and boys, AUC = .82, 95%CI [.78, .85], $p < .0001$. Focusing on girls, a cut-off score of 6.50 provided a sensitivity of .83 and a specificity of .75. Twenty-nine percent of girls ($N = 303$) scored above this cut-off. The PPV was .16 and the NPV was .99. That is, in a sample of 100 girls 29 would obtain a score over 6.50. Of these 29 children, 85% would be false positives ($N = 25$) and 15% would be true positives ($N = 4$). Of those scoring under 6.50 ($N = 71$), 70 (99%) would be true negatives and 1 (1%) would be a false negative. For boys, a cut-off score of 8.50 provided a sensitivity of .83 and specificity of .68. Thirty-eight percent of boys scored above this cut-off. The PPV was .29 and the NPV was .96. So, in a sample of 100 boys 38 would obtain a score over 8.50. Of these children, 27 (71%) would be false positives and 11 (29%) would be true positives.

Of the 62 boys scoring below the cut-off, 60 (96%) of these would be true negatives and 2 (4%) would be false negatives.

We also examined cut-off scores for children in receipt of Pupil Premium. For this group, a cut-off score of 8.50 (met by 45% of children in receipt of Pupil Premium) had a sensitivity of .85 and specificity of .63; AUC = .80, 95%CI [.75, .85], $p < .0001$. The PPV was .31 and the NPV was .96. Thus, in a sample of 100 children in receipt of Pupil Premium, 45 would score above 8.50, of whom 69% ($N = 31$) would be false positives and 31% ($N = 14$) would be true positives. Of the 55 children scoring below the cut-off, 96% ($N = 53$) of these would be true negatives and 4% ($N = 2$) would be false negatives.

In addition to these analyses on the total BESSI scores we also performed ROC analyses on the summed raw scores for each of the four BESSI factors. Table 7 shows the AUC, sensitivity and specificity of at different cut-off scores on each of the four BESSI subscales.

Discussion

This study of BESSI scores for 2106 children, including 195 with identified SEND yielded several encouraging results. First, BESSI subscale scores showed measurement invariance, indicating that meaningful comparisons could be made between children with and without identified SEND. Importantly, SEND status predicted substantial unique variance in BESSI scores, even when background effects of age, gender, ethnic minority status, family size, family income and family support were all taken into account. Specifically, compared with their 'non-SEND' peers, children with identified SEND had BESSI subscale scores that were approximately: (i) 1.5 SD higher for problems of daily living skills; (ii) 1 SD higher for problems of language/cognition and behavioral adjustment; (iii) 0.9 SD higher for problems of family support. These contrasts demonstrate the suitability of the BESSI as a screening tool for identifying children who may have SEND. A cut-off of 8.50 for BESSI total scores

showed ‘true-negative’ and ‘true-positive’ rates of 97% and 25% respectively. Finally, the cut-off score of 8.50 showed similar sensitivity and specificity among different age groups, and among children eligible for Pupil Premium, but a lower cut-off score of 6.50 was appropriate for girls. Below, we discuss these three sets of findings before outlining both study limitations and promising avenues for future research.

Children with SEND Obtain Higher BESSI Problem Ratings than their Peers

BESSI subscale scores were all significantly higher for children with SEND than for their peers. The most common problems displayed by children with SEND included problems of distractibility, unclear speech, inability to work independently, failure to understand wh-questions, and difficulty sitting still (69%). While children with SEND were more likely to show a problem on all BESSI items, these differences were especially pronounced for the child scales (Daily Living Skills, Language and Cognition, and Behavioral Adjustment). With the exception of one item (‘does not talk about fun at home’), Family Support items showed only small effects, indicating that, in general, children with SEND did not receive substantially lower levels of family support than their peers. This is reassuring because family support is an independent predictor of developmental progress (*Reference masked*), and also because it suggests that the Family Support scale shows good divergent validity.

At first glance, it appears that the large effects for two items in the Language and Cognition subscale ‘understands wh-questions’ and ‘(does not) speak clearly’ confirm the centrality of language difficulties among young children with SEND, as reported in a US Head Start preschool sample (Fantuzzo et al., 1999). However, as noted in the results section, the ‘speaks clearly’ item showed marked differential item functioning, such that comparisons at the item level may well be misleading. In particular, our analyses suggest that teachers may be unduly quick to rate a child with a statement of SEND as having problems with speaking

while erring in the opposite direction when rating children's lack of enjoyment with identifying letters.

It is also worth noting that three of the six items in the Daily Living Skills subscale also yielded large effects. This subscale was included in the BESSI in direct response to the findings from initial focus groups that emphasized the burden on teaching staff associated with child problems in daily living skills. Viewed in this context, our results highlight the need for policy and curriculum changes to help young children become more independent.

Likewise, ten of the twelve Behavioral Adjustment items showed a medium effect, including 'can play with lots of children', indicating the value of attending to social and behavioral difficulties as well as language difficulties in children with SEND. Previous research has indicated that children's behavior as well as their cognitive abilities predicts the likelihood of referral to special educational services, and that interestingly low levels of prosocial behavior are a better predictor than externalizing behaviors (Mann, McCartney, & Park, 2007). Mann et al. (2007) suggest that behavior towards peers may reveal difficulties that are masked in the more structured and supportive context of parent-child interactions. Given that peer play is widely thought to stimulate children's cognitive development (Bodrova & Leong, 1996), interventions that promote social inclusion for children with SEND may prove very valuable.

The BESSI Is Sensitive to SEND Status

From our ROC analyses, a cut-off score corresponding to problems on 8.5 or more of the 30 BESSI items showed very good sensitivity to SEND: among children with scores below this cut-off, just 3% had been identified as having SEND. The BESSI's specificity was somewhat lower, resulting in a positive predictive value of 25%, suggesting that the cut-off captures three children without identified SEND for every child with SEND. This low positive predictive value is open to at least two competing interpretations. One the one hand,

the BESSI cut-off scores may lead to a degree of over-identification, such that additional observations and assessments are needed to identify children with transient rather than long-term difficulties. Given the wide variation in normative rates of development, distinguishing between children who are 'late-starters' and those who will show persistent difficulties is an inherently challenging task. As argued in the introduction, it is probably much more important from an educational perspective to achieve sensitivity than specificity. Specifically, because the referral process is difficult, time-consuming and expensive, a low positive predictive value may simply indicate that a significant minority of children who experience significant difficulties do not yet have a statement of SEND. Taken together these findings suggest that the BESSI shows promise as an early screening tool: it is very unlikely to miss cases of SEND and provides a quick and simple means of identifying the relatively small number of children in each class who warrant more detailed evaluations.

Limitations and Future Directions

One obvious limitation of the current study is that the term 'SEND' encompasses a broad array of problems, including physical problems. It would therefore be useful for future studies to include more fine-grained comparisons of BESSI subscale scores for children with different types of SEND. Likewise, more detailed information about the children's families would be helpful to examine the family features associated with identification of SEND (e.g., parent education, Baird, et al., 2006). As described elsewhere (*Reference Masked)*, the focus groups conducted during the development of the BESSI highlighted teachers' twin concerns about (i) variability in levels of family support and (ii) the need to keep assessment measures as simple as possible. While our Family Support sub-scale was constructed in response to these dual concerns, the inclusion of more detailed measures of children's home environments would be useful in future studies. Evidence to support the potential utility of parent-implemented home therapy programmes to ameliorate speech and language

difficulties comes from a recent systematic review (Tosh, Arnott, & Scarinci, 2016); future studies of school-based influences on children's development should therefore also include measures of home-based support. Likewise, this study was restricted to formal settings of early care (i.e., nurseries and pre-schools) and so further work is needed to test the validity of the BESSI in informal care settings (e.g., child-minders). As a first step towards this goal we have shown that parental ratings of externalizing problems on Achenbach's (1992) Child Behavior Check List at age 3.5 years show robust associations with behavioural adjustment scores on the BESSI at age 5 (Darshane, 2016).

A further limitation of the current study is the lack of information about early internalizing problems or specific socio-cognitive difficulties (e.g., in empathy or perspective taking). As described elsewhere (*Reference masked)* focus groups with Foundation Years Teachers revealed widespread concerns about externalising problems, which informed the selection of items for the BESSI. In hindsight, had we included focus groups with health professionals alongside those conducted with teachers, we might have attempted to include items relating to early symptoms of anxiety or depression. However, any expansion of scope would be difficult, in that feedback from focus groups demonstrated that the BESSI's acceptability for teachers hinged on its brevity. It is, however, worth noting that children's acting out behaviors may reflect underlying problems of depression or anxiety (Polier, Vloet, Herpertz-Dahlmann, Laurens, & Hodgins, 2012). Addressing this possibility as well as the potential role of skills related to theory of mind (e.g., perspective taking, empathy) would be a valuable direction for future research. It is also worth noting that the brief nature of the BESSI enables it to be used in conjunction with other measures, including instruments that do focus on internalizing problems.

It would also be useful to gather more information about the wellbeing of teachers completing the BESSI. Studies of parent-rated questionnaires suggest that parent

characteristics influence their ratings of their child's behavior problems (Briggs-Gowan, Carter, & Schwab-Stone, 1996). For example, in a recent study of toddlers at risk of having ASD (Karp, et al., 2017), both parental wellbeing and toddlers' expressive language predicted parental concerns about their child, and parents with high levels of wellbeing expressed very few concerns about their child's behavior. Exploring whether similar effects are observed among teachers is an important step to fully understanding the meaning of BESSI scores. Similarly, a recent review suggested geographic contrasts in the prevalence of ASD, suggesting that identification of SEND may also depend on the educational policies and resources in the country or area where the child lives.

Finally, longitudinal work is needed to examine the developmental trajectories for children with and without SEND. For example, do children with SEND "catch up" with their peers, or is there a widening gap? This work could also shed light on the children who score above the 8.50 cut-off but were not identified as having SEND, by examining whether these children show similar trajectories to children with an SEND, or indeed if they are later identified as having SEND. More broadly, regularly screening is likely to be beneficial as different types of problems may emerge at different ages, and children's elevated scores at a single time point may reflect the effect of transient stressors (e.g., adjusting to parental divorce) rather than SEND. Using the BESSI throughout the year may provide teachers and parents with a simple yet reliable method of monitoring progress and informing future educational and intervention choices. These are high priorities for policy to improve the inclusivity of children's early educational experiences (Runswick-Cole, 2011).

References

Achenbach, T. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Baird, G., E, S., Pickles A, Chandler S, Loucas T, Meldrum D, & T, C. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet., 368*, 210-215. doi: 10.1016/S0140-6736(06)69041-7

Bodrova, E., & Leong, D. (1996). *Tools of the Mind: The Vygotskian Approach to Early Childhood Education*. Englewood Cliffs, New Jersey: Merrill/Prentice Hall.

Briggs-Gowan, M., Carter, A., & Schwab-Stone, M. (1996). Discrepancies among mother, child, and teacher reports: Examining the contributions of maternal depression and anxiety. *Journal of Abnormal Child Psychology, 24*, 749-765. doi: 10.1007/BF01664738

Briggs-Gowan, M., Carter, A., Skuban, E., & Horwitz, S. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old children. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 811-819. doi: 10.1097/00004583-200107000-00016

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders, 39*, 303-324. doi: 10.1080/13682820310001625589

Broomfield, J., & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomized control trial. *International Journal of Language and Communication Disorders, 46*, 628-640. doi: 10.1111/j.1460-6984.2011.00039.x

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (Second ed.). London: Guilford Press.

Byrne, B. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.

Darshane, N. (2016). *Adjustment during transition to reception and secondary school: The role of children's cognitive skills and proximal family factors.* Doctor of Philosophy, University of Cambridge, Cambridge.

Davis, T., Mayeaux, E., Fredrickson, D., Bocchini Jr., J., Jackson, R., & Murphy, P. (1994). Reading ability of parents compared with reading level of pediatric patient education materials. *Pediatrics, 93*, 460-468.

DofE. (2016). *Special educational needs in England: January 2016*. London: Department for Education.

Egger, H., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*, 313-337. doi: 10.1111/j.1469-7610.2006.01618.x

Enders, C. (2001). A Primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8*, 128-141. doi: 10.1207/S15328007SEM0801_7

Fantuzzo, J., Stoltzfus, J., Lutz, M., Hamlet, H., Balraj, V., Turner, C., & Mosca, S. (1999). An evaluation of the special needs referral process for low-income preschool children with emotional and behavioral problems. *Early Childhood Research Quarterly, 14*, 465-482. doi: 10.1016/S0885-2006(99)00030-7

Janus, M., & Offord, D. R. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science, 39*, 1-22. doi: 10.1037/cjbs2007001

Johnson, H., Carroll, J., & Bradley, L. (2017). *SEN support: a survey of schools and colleges*. London: Department for Education.

Karoly, L., Kilburn, M., & Cannon, J. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND Corporation.

Karp, E., Ibañez, L., Warren, Z., & Stone, W. (2017). Brief Report: What Drives Parental Concerns About Their 18-Month-Olds at Familial Risk for Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders, 47*, 1535-1541. doi: 10.1007/s10803-017-3060-1

Kline, R. (2012). *Principles and Practice of Structural Equation Modeling (Methodology in the Social Sciences)* (Third ed.). New York: Guilford Press.

Lavigne, J., Arend, R., Rosenbaum, D., Binns, H., Christoffel, K., & Gibbons, R. (1998). Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 1246-1254.

Lavigne, J., Gibbons, R., Christoffel, K., Arend, R., Rosenbaum, D., Binns, H., . . . Isaacs, C. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 204-214. doi: 10.1097/00004583-199602000-00014

Lumu, L., Albertyn, L., & Szabo, C. (2015). Psychiatric services for preschoolers: An emerging need. *Journal of Child and Adolescent Mental Health, 27*, 113-124. doi: 10.2989/17280583.2015.1081207

Lyon, G., Fletcher, J., Shaywitz, S., Shaywitz, B., Torgesen, J., Wood, F., . . . Olsen, R. (2002). Rethinking learning disabilities, from www.ed.excellence.net/library/special_ed_special_ed_ch12.pdf

Mandell, D., Listerud, J., Levy, S., & Pinto-Martin, J. (2002). Race differences in the age at diagnosis among medicaid-eligible children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1447-1453. doi: 10.1097/00004583-200212000-00016

Mandell, D., Novak, M., & Zubritsky, C. (2005). Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics, 116*, 1480-1486. doi: 10.1542/peds.2005-0185

Mann, E., McCartney, K., & Park, J. (2007). Preschool predictors of the need for early remedial and special education services. *Elementary School Journal, 107*, 273-285. doi: 10.1086/511707

McIntyre, L. (2008). Parent training for young children with developmental disabilities: Randomized controlled trial. *American Journal on Mental Retardation, 113*, 356-368. doi: 10.1352/2008.113:356-368

McKay, J., & Neal, J. (2009). Diagnosis and disengagement: Exploring the disjuncture between SEN policy and practice. *Journal of Research in Special Educational Needs, 9*, 164-172. doi: 10.1111/j.1471-3802.2009.01129.x

Muthén, B., & Muthén, L. (2012). *MPlus: Statistical Analysis with Latent Variables*. Los Angeles: Muthén, B & Muthén, LK.

Polier, G., Vloet, T., Herpertz-Dahlmann, B., Laurens, K., & Hodgins, S. (2012). Comorbidity of conduct disorder symptoms and internalizing problems in children: Investigating a community and a clinical sample. *European Child and Adolescent Psychiatry, 21*, 31-38.

Runswick-Cole, K. (2011). Time to end the bias towards inclusive education? *British Journal of Special Education, 38*, 112-119. doi: 10.1111/j.1467-8578.2011.00514.x

Rydz, D., Shevell, M., Majnemer, A., & Oskoui, M. (2005). Developmental screening. *Journal of Child Neurology, 20*, 4-21. doi: 10.1177/08830738050200010201

Sand, N., Silverstein, M., Glascoe, F., Gupta, V., Tonniges, T., & O'Connor, K. (2005). Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? *Pediatrics, 116*, 174-179. doi: 10.1542/peds.2004-1809

Shonkoff, J. (2003). From Neurons to Neighborhoods: Old and new challenges for developmental and behavioral pediatrics. *Journal of Developmental and Behavioral Pediatrics, 24*, 70-76.

Skovgaard, A., Olsen, E., Christiansen, E., Houmann, T., Landorph, S., Jørgensen, T., . . . Skovgaard, A. (2008). Predictors (0-10 months) of psychopathology at age 1 1/2 years - A general population study in The Copenhagen Child Cohort CCC 2000*. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49*, 553-562. doi: 10.1111/j.1469-7610.2007.01860.x

Szaniecki, E., & Barnes, J. (2016). Measurement Issues: Measures of infant mental health. *Child and Adolescent Mental Health, 21*, 64-74. doi: 10.1111/camh.12105

Tosh, R., Arnott, W., & Scarinci, N. (2016). Parent-implemented home therapy programmes for speech and language: a systematic review. *International Journal of Language and Communication Disorders, 52*, 253-269.

Table 1. *Item-Level Statistics for the BESSI*

	Whole Sample		% in Risk Category			
Item	*N*	*M (SD)*	Control	SEND	χ^2 (*1*)	OR
Behavioral Adjustment						
1 Good at waiting	2103	0.25 (0.44)	22.7	52.6	82.58	3.77
2 Calming down	2100	0.16 (0.37)	13.5	44.3	122.60	5.11
3 [R]Easily distracted	2103	0.47 (0.50)	43.6	82.0	103.79	5.87
4 [R]Easily frustrated	2094	0.26 (0.44)	22.5	57.3	111.03	4.63
5 [R]Grabs belongings	2096	0.15 (0.36)	12.9	38.0	85.74	4.15
6 [R]Often interrupts	2097	0.28 (0.45)	25.8	44.0	29.11	2.26
7 Play with children	2104	0.20 (0.45)	16.3	54.6	162.31	6.17
8 Happy to share	2097	0.16 (0.36)	13.0	42.0	110.98	4.83
9 Respectful to adults	2094	0.07 (0.25)	5.9	18.3	41.15	3.59
10 [R]Temper tantrums	2105	0.19 (0.39)	16.1	46.4	105.85	4.52
11 [R]Reprimands	2103	0.18 (0.38)	15.9	35.1	44.25	2.85
12 [R]Trouble sitting still	2094	0.35 (0.48)	31.6	68.6	105.42	4.71
Language and Cognition						
13 Speaks clearly	2104	0.27 (0.45)	21.8	81.4	315.11	15.71
14 Enjoys letters	2098	0.30 (0.46)	27.4	52.6	53.34	2.93
15 Wh-questions	2098	0.26 (0.44)	21.8	71.8	227.16	9.12
16 Recognise name	2099	0.21 (0.40)	18.8	37.9	39.67	2.64
17 1-to-1 counting	2097	0.21 (0.41)	18.4	46.1	80.98	3.78
18 Songs and rhymes	2101	0.05 (0.23)	4.6	12.9	24.18	3.09
Daily Living Skills						
19 Work independently	2106	0.29 (0.45)	24.3	73.3	206.89	8.54
20 Using scissors	2097	0.15 (0.35)	11.1	48.7	200.89	7.61
21 Help with fork	2016	0.11 (0.31)	7.9	38.4	165.78	7.23
22 Fully toilet trained	2098	0.07 (0.25)	5.4	21.6	72.17	4.82
23 [R]Appears aimless	2100	0.23 (0.42)	20.1	52.3	103.77	4.37
24 [R]Help belongings	2097	0.29 (0.42)	26.0	57.5	84.71	3.85
Family Support						
25 Receives praise	2094	0.07 (0.25)	6.2	9.9	3.89	1.66
26 Always punctual	2089	0.14 (0.34)	12.6	22.6	14.93	2.01
27 Rarely misses a day	2105	0.13 (0.33)	11.9	21.5	14.79	2.03
28 Fun at home	2101	0.29 (0.46)	25.9	63.6	121.37	5.01
29 Reads at home	1922	0.22 (0.42)	20.0	41.2	44.03	2.80
30 [R]Appears sleepy	2102	0.18 (0.38)	16.0	36.1	48.17	2.96

Table 2. *Tetrachoric Correlation Matrix for BESSI Items*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
1	-																												
2	.71	-																											
3	.77	.58	-																										
4	.73	.82	.58	-																									
5	.50	.42	.47	.53	-																								
6	.72	.80	.57	.77	.43	-																							
7	.62	.68	.62	.63	.48	.67	-																						
8	.79	.57	.86	.57	.44	.61	.61	-																					
9	.82	.69	.63	.69	.67	.72	.61	.67	-																				
10	.78	.69	.66	.66	.51	.71	.55	.70	.83	-																			
11	.73	.66	.63	.71	.54	.76	.53	.64	.76	.67	-																		
12	.78	.62	.61	.59	.37	.68	.55	.66	.64	.67	.67	-																	
13	.38	.34	.52	.29	.55	.25	.34	.47	.43	.41	.35	.22	-																
14	.45	.34	.56	.32	.38	.27	.35	.50	.44	.30	.39	.22	.54	-															
15	.48	.35	.57	.37	.53	.27	.37	.49	.46	.41	.37	.29	.78	.56	-														
16	.35	.24	.45	.27	.28	.22	.25	.34	.32	.28	.30	.20	.49	.67	.61	-													
17	.35	.27	.48	.27	.35	.24	.30	.41	.38	.29	.29	.19	.57	.69	.65	.71	-												
18	.36	.34	.47	.36	.51	.28	.45	.41	.42	.34	.36	.25	.52	.53	.51	.42	.42	-											
19	.42	.29	.39	.35	.38	.34	.19	.37	.40	.45	.52	.24	.28	.37	.39	.22	.32	.33	-										
20	.29	.35	.46	.33	.35	.31	.35	.36	.31	.33	.28	.28	.38	.33	.35	.25	.35	.41	.52	-									
21	.22	.16	.27	.16	.17	.19	.10	.26	.22	.26	.24	.21	.29	.28	.26	.21	.30	.17	.48	.42	-								
22	.19	.17	.21	.12	.25	.16	.17	.16	.17	.20	.15	.16	.31	.32	.31	.31	.29	.19	.29	.34	.69	-							
23	.34	.33	.46	.37	.58	.30	.29	.42	.47	.38	.41	.20	.68	.60	.67	.51	.55	.51	.62	.41	.30	.35	-						
24	.28	.24	.42	.24	.29	.22	.25	.36	.28	.30	.25	.23	.37	.40	.46	.41	.43	.21	.59	.38	.57	.50	.58	-					
25	.62	.46	.72	.52	.65	.38	.54	.63	.57	.54	.45	.45	.64	.62	.66	.52	.63	.56	.36	.43	.17	.24	.60	.45	-				
26	.62	.51	.61	.53	.60	.45	.48	.61	.62	.58	.52	.47	.56	.54	.65	.56	.57	.50	.36	.40	.19	.27	.55	.39	.70	-			
27	.42	.35	.51	.34	.59	.32	.35	.47	.40	.37	.34	.37	.52	.44	.50	.35	.46	.45	.29	.47	.17	.22	.49	.33	.66	.52	-		
28	.32	.30	.31	.31	.44	.18	.39	.29	.33	.33	.22	.22	.56	.27	.58	.41	.38	.33	.24	.27	.26	.30	.41	.30	.46	.53	.28	-	
29	.59	.48	.66	.43	.44	.42	.52	.59	.54	.55	.41	.50	.46	.48	.47	.44	.43	.43	.24	.41	.18	.25	.40	.33	.61	.63	.50	.38	-
30	.30	.31	.32	.29	.41	.25	.27	.22	.33	.33	.34	.22	.44	.46	.42	.49	.44	.30	.09	.26	.10	.26	.36	.22	.43	.53	.26	.34	.45

Note. Numbers 1 to 30 refer to the Items in Table 1.

Table 3. *Model Fit Statistics*

	Model	χ^2	*df*	CFI	TLI	RMSEA (90% CI)	$\Delta\chi^2$
	Measurement CFA (Whole Sample)						
1	4 factor model	2433.29	399	0.951	0.947	0.049 (0.047, 0.051)	-
	Multiple Groups CFA (Whole Sample)						
	Tests of Measurement Invariance						
2	Equal Form	2317.40	798	0.952	0.948	0.043 (0.041, 0.045)	-
3	Equal Form and Equal Loadings	2160.03	824	0.958	0.955	0.039 (0.037, 0.041)	34.591
4	Equal Form, Loadings, and Thresholds	2268.24	850	0.955	0.954	0.040 (0.038, 0.042)	195.55
5	Equal Form, Loadings, and Thresholds (Item Residuals Free)	2393.66	820	0.950	0.947	0.043 (0.041, 0.045)	-
6	Equal Form, Loadings, and Thresholds (Item Residuals Free; Two Thresholds Free)	2345.58	818	0.952	0.949	0.042 (0.040, 0.044)	-
7	Equal Form, Loadings, Thresholds, Item Residuals (Two Thresholds Free)	2231.52	848	0.956	0.955	0.039 (0.037, 0.041)	52.784*
	Tests of Population Heterogeneity						
8	Equal Factor Variance: Behavior Adjustment	2236.81	849	0.956	0.955	0.039 (0.037, 0.041)	9.216
9	Equal Factor Variance: Behavior Adjustment, Language and Cognition	2250.44	850	0.956	0.955	0.040 (0.038, 0.042)	13.10*
10	Equal Factor Variance: Behavior Adjustment, Language and Cognition, Daily Living Skills	2245.55	851	0.956	0.955	0.039 (0.037, 0.041)	1.410
11	Equal Factor Variance: Behavior Adjustment, Language and Cognition, Daily Living Skills, Family Support	2241.91	852	0.956	0.955	0.039 (0.037, 0.041)	1.388
12	Partial Invariance Model, Equal Factor Variances, Equal Factor Mean: Behavior Adjustment	2764.56	853	0.940	0.938	0.046 (0.044, 0.048)	191.759*
13	Partial Invariance Model, Equal Factor Variances, Equal Factor Mean: Language and Cognition	2409.04	853	0.951	0.950	0.042 (0.040, 0.044)	164.944*
14	Partial Invariance Model, Equal Factor Variances, Equal Factor Mean: Daily Living Skills	2620.64	853	0.944	0.943	0.044 (0.042, 0.046)	306.767*
15	Partial Invariance Model, Equal Factor Variances, Equal Factor Mean: Family Support	2361.355	853	0.952	0.951	0.041 (0.039, 0.043)	96.054*
	Multilevel CFA						
16	4 factors, 2 levels, factor loadings freely estimated	2003.71	798	0.933	0.927	0.027	-

17	4 factors, 2 levels, factor loadings freely estimated, between-level residual for item 2 set at 0.	2004.92	799	0.933	0.927	0.027	-
18	4 factors, 2 levels, factor loadings equal across levels; between-level residual for item 2 set at 0.	1706.82	825	0.951	0.948	0.023	-
	Multilevel SEM						
19	4 factors (at within-level) regressed onto SEND Status, Age, Gender, Pupil Premium Status, Ethnic Minority Status, Sibling Status.	1849.57	979	0.940	0.936	0.022	-
20	3 factors (at within-level) regressed onto SEND Status, Age, Gender Pupil Premium Status, Ethnic Minority Status, Sibling Status, and Family Support Latent Factor Scores.	2048.14	992	0.943	0.939	0.023	-

Note. 90% CIs for the RMSEA cannot be computed for multi-level models in Mplus. *$p < .0028$. Chi-square difference tests are not reported for non-nested models.

Table 4. *Unstandardized and standardized parameter estimates for multilevel CFA (Loadings freely estimated across levels).*

Factor	Item	Single Level			ICC	Within			Between		
		Est.	SE	Std.		Est.	SE	Std.	Est.	SE	Std.
BA	1 Good at waiting	1.00	-	.90	.10	1.00	-	.91	1.00	-	.87
	2 Calming down	0.91	0.02	.82	.13	0.61	0.05	.80	0.72	0.15	.80
	3 [R]Easily distracted	0.99	0.02	.90	.09	1.04	0.11	.92	0.86	0.20	.78
	4 [R]Easily frustrated	0.91	0.02	.82	.15	0.60	0.05	.80	0.81	0.16	.83
	5 [R]Grabs belongings	0.80	0.03	.72	.23	0.47	0.04	.72	0.88	0.20	.80
	6 [R]Often interrupts	0.89	0.02	.80	.12	0.67	0.05	.83	0.58	0.14	.62
	7 Play with children	0.81	0.02	.73	.19	0.49	0.04	.74	0.69	0.16	.69
	8 Happy to share	0.96	0.02	.87	.09	0.82	0.07	.88	0.81	0.18	.84
	9 Respectful to adults	0.97	0.02	.87	.16	0.90	0.07	.89	1.13	0.21	.83
	10 [R]Temper tantrums	0.93	0.02	.84	.12	0.72	0.07	.85	0.77	0.14	.80
	11 [R]Reprimands	0.90	0.03	.81	.13	0.64	0.06	.82	0.62	0.17	.65
	12 [R]Trouble sitting still	0.82	0.02	.74	.15	0.51	0.04	.75	0.64	0.13	.73
LC	13 Speaks clearly	1.00	-	.82	.17	1.00	-	.80	1.00	-	1.00
	14 Enjoys letters	0.96	0.03	.79	.40	1.15	0.11	.83	1.39	0.32	.72
	15 Wh-questions	1.06	0.03	.87	.30	1.47	0.15	.89	1.66	0.27	.88
	16 Recognise name	0.87	0.03	.72	.41	0.99	0.11	.79	0.76	0.27	.42
	17 1-to-1 counting	0.94	0.03	.78	.27	0.93	0.08	.77	0.89	0.21	.69
	18 Songs and rhymes	0.84	0.05	.69	.27	0.75	0.09	.70	0.79	0.22	.70
DLS	19 Work independently	1.00	-	.88	.33	1.00	-	.91	1.00	-	.69
	20 Using scissors	0.96	0.03	.84	.11	0.86	0.09	.88	1.42	0.31	.70
	21 Help with fork	0.76	0.03	.68	.17	0.41	0.04	.67	0.52	0.10	.69
	22 Fully toilet trained	0.63	0.04	.56	.23	0.37	0.04	.63	0.67	0.26	.47
	23 [R]Appears aimless	0.85	0.03	.74	.26	0.53	0.05	.76	0.81	0.21	.70
	24 [R]Help belongings	0.61	0.05	.53	.28	0.27	0.04	.51	0.40	0.13	.61
FS	25 Receives praise	1.00	-	.70	.18	1.00	-	.70	1.00	-	.68
	26 Always punctual	0.95	0.07	.67	.35	0.89	0.15	.66	0.53	0.19	.76
	27 Rarely misses a day	0.77	0.06	.54	.16	0.61	0.09	.51	0.48	0.16	.61
	28 Fun at home	0.76	0.07	.53	.42	0.58	0.08	.50	0.60	0.20	.63
	29 Reads at home	1.32	0.08	.92	.25	2.32	0.54	.92	1.85	0.60	.84
	30 [R]Appears sleepy	0.99	0.07	.70	.16	0.97	0.14	.69	0.96	0.30	.74

Note. BA = Behavioral Adjustment. LC = Language and Cognition. DLS = Daily Living Skills. FS = Family Support.

Table 5. *Multi-level Structural Equation Model Parameter Estimates (Within-Level).*

	Behavioral Adjustment			Language and Cognition			Daily Living Skills			Family Support		
Predictor	**Est.**	**S.E.**	**Std. Est.**	**Est.**	**S.E.**	**Std. Est.**	**Est.**	**S.E.**	**Std. Est.**	**Est.**	**S.E.**	**Std. Est.**
Model 11												
SEND Status	2.08	0.22	.27*	1.38	0.12	.33*	3.17	0.31	.40*	1.06	0.19	.23*
Age	-0.02	0.01	-.06	-0.07	0.01	-.42*	-0.07	0.01	-.23*	-0.03	0.01	-.12
Gender	1.04	0.14	.23*	0.45	0.07	.18*	1.25	0.14	.27*	0.35	0.10	.13*
Pupil Premium Status	0.41	0.14	.08	0.40	0.08	.14*	0.57	0.15	.11*	1.10	0.16	.35*
Ethnic Minority Status	0.06	0.17	.01	0.51	0.09	.17*	0.23	0.19	.04	0.46	0.13	.13*
Older Sibling	-0.46	0.13	-.10*	0.07	0.07	.03	-0.39	0.14	-.08	0.26	0.09	.09
Model 12												
SEND Status	1.23	0.19	.16*	0.78	0.10	.17*	1.73	0.20	.24*	-	-	-
Age	-0.02	0.01	-.07	-0.08	0.01	-.45*	-0.06	0.01	-.22*	-	-	-
Gender	1.08	0.13	.25*	0.58	0.08	.22*	1.33	0.13	.32*	-	-	-
Pupil Premium Status	-0.38	0.13	-.07	-0.23	0.08	-.08	-0.45	0.14	-.09	-	-	-
Ethnic Minority Status	0.12	0.16	.02	0.47	0.09	.14*	0.29	0.18	.06	-	-	-
Older Sibling	-0.29	0.12	-.07	0.17	0.07	.06	-0.17	0.13	-.04	-	-	-
Family Support	0.99	0.12	.47*	0.76	0.11	.60*	1.27	0.17	.63*	-	-	-

Note. *$p < .001$.

Table 6. *Sensitivities and Specificities for Various Total Scores on the BESSI for Predicting SEND status*.

BESSI Raw	**Whole Sample**		**Boys**		**Girls**		**2.50 – 3.49 Years**		**3.50 – 4.49 Years**		**4.50 – 5.50 Years**	
Total Score	**Sens.**	**Spec.**	**Sens.**	**Spec.**	**Sens.**	**Spec.**	**Sens.**	**Spec.**	**Sens.**	**Spec.**	**Sens.**	**Spec.**
3.50	.93	.50	.95	.41	.90	.58	.94	.36	.94	.48	.93	.56
4.50	.92	.56	.94	.47	.88	.64	.94	.45	.92	.54	.91	.62
5.50	.87	.62	.89	.53	.83	.71	.88	.52	.88	.60	.85	.67
6.50	.86	.66	.88	.57	.83	.75	.82	.57	.87	.65	.85	.70
7.50	.82	.71	.83	.63	.77	.78	.82	.63	.83	.68	.79	.76
8.50	.80	.75	.83	.68	.70	.81	.82	.68	.79	.73	.77	.80
9.50	.75	.78	.79	.72	.65	.84	.82	.73	.75	.75	.72	.82

Note. Sens. = Sensitivity. Spec. = Specificity.

Table 7. *AUCs, Sensitivities and Specificities for the Summed BESSI Subscale Scores for Predicting SEND Status in the Whole Sample*.

	Behavior Adjustment		**Language and Cognition**		**Daily Living Skills**		**Family Support**	
	AUC = .78, 95%CI [.75, .81]		AUC = .80, 95%CI [.77, .83]		AUC = .82, 95%CI [.79, .85]		AUC = .71, 95%CI [.67, .74]	
Total Raw Score	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
1.5	.86	.55	.81	.70	.78	.75	.51	.53
2.5	.79	.66	.62	.82	.60	.86	.33	.76
3.5	.69	.74	.40	.89	.39	.94	.20	.88

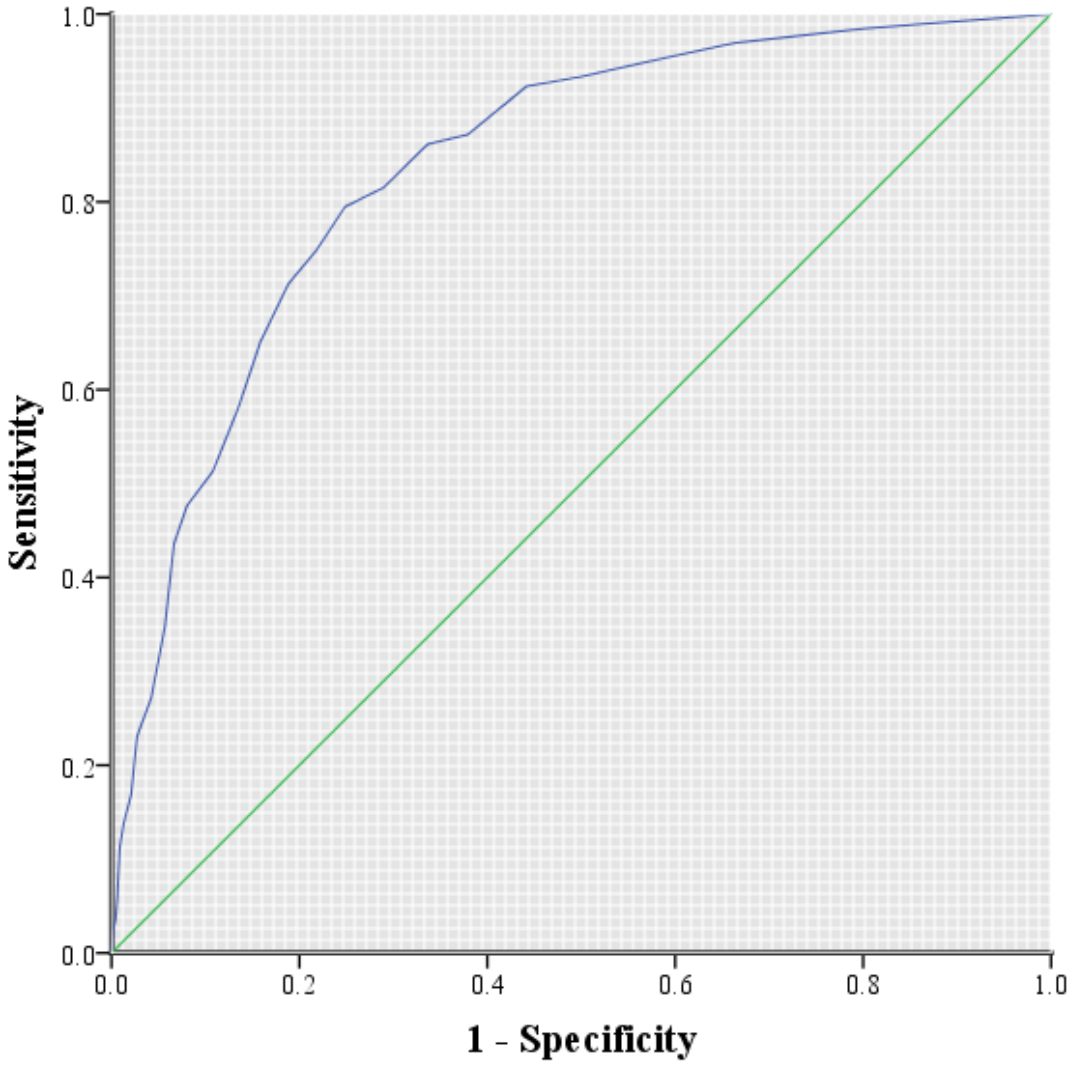

Figure 1. ROC curve for BESSI total raw scores.