Validation of the Wider Outcomes Survey for Teachers (WOST): A measure for assessing the behaviour, relationships and exposure to bullying of children and young people with Special Educational Needs or Disabilities (SEND)

Michael Wigelsworth[1], Jeremy Oldfield[2] and Neil Humphrey[1]

[1]*School of Education, University of Manchester, UK*

[2] *Department of Psychology, Edge Hill University, UK*

Address for correspondence:

Dr. Michael Wigelsworth

Educational Support and Inclusion

School of Education

University of Manchester

Oxford Road

Manchester

M13 9PL

United Kingdom

Tel: (00 44) (0)161 306 1763

Email: michael.wigelsworth@manchester.ac.uk

1

Validation of the Wider Outcomes Survey for Teachers (WOST): A measure for assessing the behaviour, relationships and exposure to bullying of children and young people with Special Educational Needs or Disabilities (SEND)

Abstract

The Wider Outcomes Survey for Teachers (WOST) is a teacher informant-report questionnaire developed to aid the assessment of behaviour difficulties, quality of relationships and exposure to bullying among students identified with special educational needs and disabilities (SEND) (Humphrey & Squires, 2011). This study examines the psychometric properties of the WOST in a validation sample representing 6164 students with SEND (mean age 12 years) drawn from 481 primary and secondary schools across England. Results showed favourable internal consistency using Cronbach's Alpha and acceptable model fit using Confirmatory Analysis (CFA), both of which were invariant to broad categorisations of SEND. Practical utility and construct validity were also established by testing two theoretically derived hypotheses. The measure is therefore tentatively supported as a useful tool for assessing the wider outcomes of students with SEND.

Keywords: Behaviour, bullying, relationships, measurement, assessment, special educational needs and disabilities, psychometric properties

Introduction

Establishing effective systems and practices to support learners with special educational needs and disabilities (SEND) is a major policy priority for many countries. Indeed, in recent years there have been developments in policy in a number of countries, including the USA (Individuals with Disabilities Education Improvement Act, 2004, see Wright, 2004), Australia (Department of Education and Training, 2004) and England (Department for Education and Skills, 2004) with the common aim of reforming existing systems in order to promote inclusion (Ferguson, 2008).

A key part of these reforms has been the move to ensure that students with SEND are educated alongside their non-disabled peers in mainstream school settings wherever possible (Department for Education and Employment (DfEE), 1997; Department for Education and Skills (DfES), 2001; DfES. 2004). Currently, students are considered as requiring support in one (or more) of four broad areas of need;

- Communication and interaction

- Cognition and learning

- Behavioural, emotional, and social development

- Sensory and/or physical

(DfES, 2001)

Although these broad categories are further subdivided into arguably more definitive areas of need (e.g. Autistic spectrum disorder, physical disability, multi-sensory impairment, etc), there is still considerable debate how SEND is defined and understood (Sebba, Peacock, DiFinizio, & Johnson, 2011; Williams, Lamb, Norwich,

& Peterson, 2009). This is because there are complex issues such as overlapping definitions, co-morbidity, variations in assessment criteria, and changes in need dependent on context and time (Riddick, 2012). Therefore, although there is consistency in the terminology used (e.g. all schools use this code of practice), this can mask an inconsistency in the way the terminology is used in practice (Mooney, Owen, & Statham, 2008; Porter, Daniels, Feller, & Georgeson, 2009), Indicating that to ensure accurate representation, the broader consideration of terminology should apply to all students coming under the umbrella of SEND.

A further concern is a lack of effective monitoring of the outcomes of inclusive practice in schools. Although learners' academic progress is typically used as an indication of the success of attempts to support those with SEND, truly inclusive practice requires a focus on the full range of school activities (Farrell, 2000), such as building social relationships and mutual respect from peers. Effective monitoring of students (by both academics and practitioners) identified as having SEND is especially important as research in this area has indicated that for these students, experience of school is often marked by higher levels of bullying, social isolation and peer rejection, and poorer behaviour when compared with their non-SEND peers (Avramidis, 2010; Wiener & Mak, 2009; Van Roekel, Scholte, & Didden, 2010).

For instance, students identified with specific language impairment (SLI) (Conti-Ramsden & Botting, 2004), learning difficulties (Mishna, 2003) or ASD (Humphrey & Symes, 2010) have been demonstrated to be at risk of increased exposure to bullying and lower quality or numbers of friendships. There are multiple and overlapping pathways to explain such findings for instance increased vulnerability

and lower levels of social support, however, authors have noted that this body of work is yet to be unified under the common factor of being identified with SEND (van Roekel et al., 2010), although a review by Rose, Monda-Amaya, and Espelage (2010) is a move in this direction. This is important as there is a suggestion that the identification of SEND itself is a potential risk factor for negative psychosocial outcomes such as lower self esteem (Taylor, Hume, & Welsh, 2010) and difficulties in building positive relationships (Frostad & Pijl, 2007) in addition to the specific nature of the need or difficulty. However, there is a clear need for further research in this area.

Given the lack of research into outcome monitoring for students identified with SEND and the potentially harmful effects of negative school experiences, a challenge has been issued by researchers to establish effective means by which the experiences of students can be assessed (Tetler & Baltzer, 2011), specifically the "*quality of the school experience and about how far [pupils identified with SEND] are helped to learn, achieve and participate fully in the life of the school"* (DfES, 2004, p. 12). Successful monitoring is arguably a necessary component in examining any effects of interventions for SEND students (i.e. the WOST measure was developed in order to assess the effects of a SEND intervention 'Achievement for All' – Humphrey and Squires, 2011) and can also help address some of the current gaps in the literature highlighted in the preceding review, both of which, in turn, have positive implications for the improvement of school practice.

Measurement of wider outcomes for learners with SEND

Although there are a number of established measures that tap 'wider outcomes' (i.e. measures that are not directly associated with attainment, such as bullying or positive relationships) of the kind noted above, their utility in monitoring school experience for students identified with SEND are limited in two key respects.

First, given obvious limitations in accessing large numbers of students identified with SEND, validation histories are most often based on normative or non-SEND school populations – known as 'analogue' samples. This is done either through ignoring a child's SEND status (for instance, in the development of the Social Bullying Involvement scales (Fitzpatrick & Bussey, 2011) and in several studies of the Olweus Bully/Victim questionnaire (Kyriakides, Kaloyirou, & Lindsay, 2006; Lee & Cornell, 2009)) meaning that there is no consideration of SEND in sampling or recruitment, or by screening for concurrent behaviours or issues. For instance, both Solheim, Berg-Nielsen, & Wichstrom (2011) and Koomen, Verschueren, van Schooten, Jak, & Pianta (2012) assessed student-teacher relationships alongside results from the Strengths and Difficulties questionnaire (Goodman, 1997).

Second, for behaviour measures that *have* been validated with SEND populations, these are typically used as screening tools in order to identify measureable behaviours for the purpose of identification or diagnoses, often featuring a clinical history, rather than the monitoring of behaviour itself. Notable examples include the Achenbach (Achenbach, 2001), the behavioural assessment system for children (BASC) (Kamphaus, 2004), and in the case of the Disruptive Behaviour Disorder Scale (DBDS), this has been validated on students specifically who fulfil the DSM-IV criteria for ADHD, ODD or CD (Van Eck, Finney, & Evans, 2010).

The authors of the paper acknowledge that such limitations are, in many cases, either unavoidable and/or acceptable limitations in many studies. For instance, a large body of research is interested in the mainstream, or majority of students without SEND. Second, pragmatic difficulties often limit the size of samples obtainable in studies focusing on SEND as large numbers of students are required, and given the small numbers per school relative to the analogue sample, this is often beyond the resources of the researcher. This is a particular difficulty given the heterogeneous nature of SEND, as there many diverse needs and circumstances that require sampling. However, given the context of negative school experience in inclusive practice raised in this review, tools that have been validated on mainstream pupils should not necessarily be inferred to be as valid for a SEND sample.

Although some papers are beginning to emerge (e.g. Koster, Minnaert, Nakken, Pijl, & van Houten, 2011) that consider these points, there is an obvious need for further research, especially given the potentially large SEND 'population' worldwide (estimated at approximately 1.67 million (one fifth) of the school population in England alone - DfE, 2011). The aim of the current study, therefore, is to advance the field via the publication of the first measurement tool designed specifically to assess wider outcomes (specifically, behaviour, bullying and positive relationships) among students with SEND, using a large, representative validation sample. The Wider Outcomes Survey for Teachers (WOST) was developed during a large-scale, national evaluation of an educational initiative that targeted learners with SEND (Humphrey and Squires, 2011). In this paper we describe the development of the WOST and present analyses pertaining to its psychometric properties, drawing upon

Terwee, Bot, de Boer et. al's, (2007) proposed validation criteria. In addition, we test the construct validity of the measure generating theoretically derived hypotheses, consistent with the concepts being measured by the tool, specifically;

- *behaviour difficulties and bullying are positively correlated, and that in turn they are both negatively correlated with positive relationships*

- *WOST subscales will discriminate between students with behavioural, emotional, social difficulties (BESD), Autism spectrum disorders (ASD) (a subset of the communication and interaction category) and those with other SEND.*

Method

Participants

The validation sample consisted of 6164 students with SEND drawn from a nationally representative, geographically diverse sample of 198 primary and 83 secondary schools across England in mid 2010.

As noted earlier, schools and students were recruited as part of large-scale, national evaluation of an educational initiative that targeted learners with SEND (Humphrey & Squires, 2011). Students were drawn from years 1, 5, 7 & 10 (aged 5-6, 9-10, 11-12 and 14-15 respectively), were all identified as having SEND, and were in receipt of special educational provision at the School Action (teachers notice a student's difficulties in one or more areas and adjust their teaching accordingly), School Action Plus (external involvement is sought, for example from an educational psychologist) or Statement of Special Educational Needs (following a statutory assessment, a

legal document is constructed that outlines the student's needs and how they may best be met; this typically secures additional financial resources that can be used to provide more intensive support than would typically be available). No other inclusion or exclusion criteria were applied.

The characteristics of the validation sample compare favourably to national SEND trends and are shown in Table 1.

Table 1

Characteristics of Sample and Comparison with National Averages of pupils identified with SEND (school action plus and statement)

Instrument

The WOST is intended to measure the teacher's perception of a student's wider outcomes at school in three key areas: behaviour difficulties (6 items), relationships with other people (7 items) and experience of bullying (7 items), making a total of 20 items. To assess behaviour and experience of bullying, teachers indicate the frequency of a series of behaviours (e.g. "*The pupil cheats and tells lies*" / "*the pupil is picked on by other children*") using a four point Likert scale (*never*= 0 / *rarely*= 1 / *sometimes*= 2 / *often*= 3). A similar scale is used to assess the quality of the pupil's relationships with other people (e.g. "*the pupil has at least one good friend*") (*strongly disagree* =0 / *disagree*= 1 / *agree*= 2 / *strongly agree*=3). The WOST is calculated by taking the item average for each domain (minimum score = 0, maximum score = 3). A copy of the instrument is shown in appendix 1.

Scale development

In this section we outline the process through which the WOST was developed. This is in the interests of transparency, but also serves to highlight content validity or, "the extent to which the concepts of interest are comprehensively represented by items in the questionnaire" (Terwee et al., 2007, p.35). It is suggested that authors should provide a clear description of the following:

1. *Measurement aim of the questionnaire*. The primary measurement aim of the WOST was evaluative, as we sought to measure to develop a tool that could be used to monitor wider outcomes in schools.

2. *Target population*. The population for which the survey was developed was (by proxy) students and young people with SEND.

3. *Concepts*. The WOST was designed to measure behaviour, bullying and positive relationships.

4. *Item selection and item reduction*. Items were generated and selected using a combination of three methods: (i) discussion and 'brainstorming' among members of the research team, (ii) reference to existing measures (e.g. the 'conduct problems' subscale of the Strengths and Difficulties Questionnaire for the 'behaviour' domain), and (iv) reference to the broader research literature pertaining to each of the concepts of interest. The surveys were piloted in autumn 2009. Basic psychometric analysis suggested the surveys were fit for purpose and so they were used in their entirety with the validation sample, with item reduction occurring largely *post‑hoc*. The initial version of the WOST contained 28 items (9 behaviour, 9 bullying and 10 positive relationships). Item reduction techniques were applied and individual items were eliminated if, for example, they were regularly skipped by a large proportion of participants, or if their removal improved one or more measurement properties (e.g.

internal consistency). The final version of the WOST contains 20 items (6 behaviour, 7 bullying and 7 positive relationships).

5.   *Item interpretability*. Items were carefully written to avoid unnecessary jargon and technical language. Interpretability of items relating to concepts with equivocal meanings (e.g. bullying) was facilitated with the use of clear definitions where appropriate.

Procedure

Following the completion of standard ethical protocols (e.g. parental opt-out consent), participating schools were provided with a link to a secure, password-protected website that housed the survey. For students in primary schools, the class teacher completed the WOST; in secondary schools, it was typically completed by their form tutor.

Data was entered into SPSS (version 16) and for the CFA, AMOS (version 7) was used.

Results

Data screening

Of the original 9136 cases, 3002 cases were identified as incomplete i.e. missing raw scores from which to calculate mean values. Tabulated pattern analysis showed that the pattern of missing scores was spread across every item and ranged between 0.7% and 1.9% of total responses. No discernable pattern was detected, as

missing scores were distributed across different types of SEND, age group and gender, and therefore were safely removed list wise from the final analysis.

Prior to conducting inferential analyses, the robustness and quality of data were examined.

First, cases were screened for outliers. Univariate statistics and Mahalanobis distances identified 31 cases as extreme, and as there was no discernible link between the cases, these were removed. Second, univariate normality was assessed through examining skewness and kurtosis. Using cut-offs of skewness >2.0 and kurtosis >7.0 (Curran, West, & Finch, 1996), data were considered within parameters for univariate normality. However, it should be noted that for the domains of behaviour and bullying, there was a general trend towards a positive skewness, indicating a partial floor effect. Third, multi-colinearity was examined using bivariate correlations. Relationships between domains did not exceed r = .618.

This resulted in the final sample of 6164 cases, far in excess of recommended sample sizes for CFA (Comfrey & Lee, 1992, Tabachnick & Fidell, 2007) and the other analyses that follow, even when accounting for the heterogeneity of the sample.

Descriptive statistics and internal consistency

Table 2 shows the means, standard deviations and internal consistency co-efficients (Cronbach's alpha) for each of the broad categories of SEN. As seen from table 2,

the Cronbach's alpha values in the current study range between .87 and .92, indicating high internal consistency across domains that is invariant to SEND classification. These values greatly exceed the standard threshold of 0.7 for attitudinal and mental health measures (Bland & Altman 1997; Henson, 2001).

Table 2. Descriptive Statistics and Reliability by SEND Classification

To further examine the internal consistency of the WOST, its hypothesized structure was assessed in a series of CFAs (for the scale as a whole and for each major classification of SEND type). In selecting appropriate fit indices, sample size was an important consideration as the traditional measure of chi-square and adjusted goodness of fit are greatly affected by this (Hu and Bentler, 1998). For the current study, chi-square was supplemented with the comparative fit index (CFI), the Tucker-Lewis coefficient (TLI), root mean square residual (RMR), and root mean square error of approximation (RMSEA). Ideal fit indices were created by examining the indices reported in a range of similar inventories, in order to estimate what is considered 'good fit' in the measurement of psycho-social outcomes (Fitzpatrick & Bussey, 2011; Georgiou, 2008; Hatami, Motamed, & Ashrafzadeh, 2009; Kim & Kamphaus, 2010; Ryser, Campbell, & Miller, 2010). The results of the CFAs are shown in table 3 and represented graphically in figure 1.

Table 3. Fit Indices for Models by SEND Classification

Figure 1. Diagram of factor loadings (standardised estimates)

Given the large sample sizes involved, x^2 is not used to suggest model fit. RMSEA is considered a more accurate indicator as it is not highly sensitive to sample size. The RMSEA of all the CFA models suggests a model close to ideal fit (Browne & Cudeck, 1992). An assessment of acceptable fit is consistent with the other indices, as CFI, TLI, AGFI and RMR are all approaching ideal fit thresholds. Of note is the consistency of the various fit indices across the broad classifications of SEND. This suggests that the WOST structure is invariant across different types of SEND.

Construct validity

In order to assess the construct validity of the WOST, theoretically derived hypotheses were generated that were consistent with the concepts being measured by the tool (Terwee et al, 2007). Previous research has suggested that students with behaviour problems are also likely to experience bullying, and vice-versa. Likewise, students regarded as having strong positive relationships in school are not likely to experience bullying and/or behaviour problems. In light of this*, we predicted that behaviour difficulties and bulling would be positively correlated, and that in turn they would both be negatively correlated with positive relationships.* Bi-variate correlations indicated moderate, statistically significant relationships between the subscales that were consistent with these predictions. Specifically these were behaviour – bullying (r = 0.618, p <0.01), bullying-positive relationships (r = -0.481, p < 0.01), and positive relationships – behaviour (r = -0.565, p <0.01).

A measure such as the WOST should also be able to discriminate between groups of learners in theoretically plausible ways. As such, we predicted that *WOST*

subscales would discriminate between students with behavioural, emotional, social difficulties (BESD), Autism spectrum disorders (ASD) (a subset of the communication and interaction category) and those with other SEND.

This hypothesis provided a test of the discriminative validity of the WOST as SEND theory and research suggests that pupils with BESD and ASD are more likely to experience negative outcomes in relation to behaviour, bullying and positive relationships than other pupils with SEND (Avramidis, 2010; Humphrey & Symes, 2010). A series of one-way ANOVAs demonstrated that the WOST was able to discriminate between these groups for each of the subscales, specifically behaviour ($F(2,6381) = 595.52$, $p<0.001$, $\eta^2 = 0.157$), bullying ($F(2,6381) = 161.03$, $p<0.001$, $\eta^2 = 0.042$) and positive relationships ($F(2,6381) = 320.49$, $p<0.001$, $\eta^2 = 0.091$). Accompanying post-hoc effect size analyses are shown in table 4.

Table 4. Discriminative validity by SEND classification

Discussion

The purpose of the study was to examine the psychometric properties of the recently developed Wider Outcomes Survey for Teachers (WOST) in a validation sample of students identified with SEND. This included assessment of the scale's content validity, internal consistency and model fit, and finally construct validity.

Results showed that internal consistency was high for each of the three domains (behaviour, relationships and bullying) and invariant across the broad categorisations

of SEND used in the study.   A confirmatory factor analysis showed that after accounting for sample size, indices indicated acceptable levels of overall fit. Although measures were lower than 'ideal' thresholds, they were consistent with comparable tools measuring similar domains (Tarshis & Huffman, 2007; Fitzpatrick & Bussey, 2011; Koster et al, 2011; Koomen et al, 2012) (this is especially true given the heterogeneity of the sample under investigation -see strengths and limitations). Despite this high level of variation within the sample, there was also invariance of model fit across the broad categories of SEND (behavioural, emotional, and social difficulties, cognition and learning, communication and interaction, and sensory and/or physical difficulties). This suggests that at a super-ordinate level, the WOST is a consistent measure of wider outcomes for students identified with SEND. The three subscales of the instrument related to one another in a manner that was consistent with expectations, enhancing construct validity. Finally, the WOST was also able to discriminate between particular groups of learners (those with BESD, ASD and other SEND) in a way that was inline with previous research and our a-priori predictions, further supporting its construct validity.

Limitations of the current study

Although a key strength of the study is the large and nationally representative sample, there are several limiting factors, suggesting caution in interpreting the results is warranted. First, as mentioned at the beginning of this article, there is still debate regarding definition and use of the term SEND (Mooney et al., 2008; Porter et al., 2009; Sebba et al., 2011; Williams et al., 2009). Whether a student is identified as having SEND is confounded by a range of issues including, variations in

assessment criteria, the resources available in individual schools, and a contentious socio-political history regarding the use of identification and categorisation of needs (Riddick, 2012). However, regardless of the myriad of issues surrounding the accuracy and validity of identifying SEND, the sample does represent those currently being monitored by schools, and who have effectively been labelled.

A second issue is the accuracy of teacher reports of 'experience of bullying' and 'quality of relationships'. Although teachers were selected as respondents on the basis of the time spent with the pupil and the use of their collective experience with other students (an aspect missing from parental responses), several studies have noted that teachers tend towards an overestimation of social participation (Nabuzoka & Smith, 1993) and an underestimation of peer victimisation (Martlew & Hodson, 1991). This highlights an important need to consider multiple perspectives when assessing psychosocial measures (Dickson, Emerson & Hatton, 2005; Wigelsworth, Humphrey, Lendrum, & Kalambouka, 2010), and although a pupil dataset was not used in the wider evaluation from which this data was sourced, a parental version of the WOST is currently being prepared for publication.

A final consideration is the cultural transferability of the concepts measured. Although inclusion is a global concern, issues of behaviour, bullying and relationships are culturally bound. This is especially true for bullying, as this varies across countries, cultures, and languages (Smith, Cowie, Olafsson, & Liefooghe, 2002). Therefore, although the WOST has a potentially global use, consideration of cultural norms and validation alongside other instruments is required.

Directions for future research

The issues highlighted above offer several avenues for future development. First, there is an opportunity to conduct further psychometric testing of the WOST by examining test-retest reliability and to assess the tool alongside other criterion measures, allowing further validation of the instrument itself. Second, the WOST potentially represents a useful addition alongside the current measures available for research, with a special focus on SEND, given its validation and highlighted limitations of similar measures (see literature review). For instance, the WOST may be of interest to researchers as part of school based evaluation designs (see Humphrey and Squires, (2012) as an example) or used alongside similar psychometrically validated instruments in order to explore the variety of influences on children's wider experiences at school. Third, there is an opportunity to use the WOST in a wider context, for instance, examining different school contexts (e.g. pupil referral units) or potential further validation outside of a UK context. In summary the WOST represents a partial answer to the challenge set by (Tetler & Baltzer, (2011), to establish effective means by which the experiences of students with SEND can be assessed, which may be a potentially valuable contribution given the current controversies and lack of research in this field.

Declaration of Conflicting Interests

Funding

References

Achenbach, T. (2001). *Manual for ASEBA school age forms and profiles*. Burlington, VT: University of Vermont, Research Centre for Children, Youth and Families.

Avramidis, E. (2010). Social relationships of pupils with special educational needs in the mainstream primary class: peer group membership and peer-assessed social behaviour. *European Journal of Special Needs Education*, *25*, 413-429.

Bland, M. & Altman, D. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal, 37*, 527

Browne, M., & Cudeck, R. (1992). Alternative ways to assessing model fit. *Sociological Methods & Research*, *21*, 230-258.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Comfrey, A.l., & Lee, H.B. (1992). *A first course in factor analysis (2nd Ed).* Hillsdale, NJ: Erlbaum

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research*, *47*, 145-61.

Curran, P., West, S., Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods. 1, 16-29*

Department for Education (DfE) (2011). *Special Educational Needs Information Act*. London: Department for Education.

Department for Education and Employment (DfEE) (1997) *Excellence for all children: Meeting special educational needs*. London: Department for Education and Employment.

Department for Education and Skills (DfES) (2001). *Inclusive schooling: children with special educational needs*. London: Department for Education and Skills.

Department for Education and Skills (DfES) (2004). *Removing barriers to achievement*. London: Department for Education and Skills.

Department of Education (2010). *Children with Special Educational Needs 2010: An Analysis*. London: Department for Education.

Department of Education and Training. (2004). *Pathways to the future: A report of the review of the educational services for students with disabilities in Government schools*. Perth: Department of Education and Training.

Dickson, K. Emerson, E. Hatton, C. (2005). Self-reported anti-social behaviour: prevalence and risk factors amongst adolescents with and without intellectual disability. *Journal of Intellectual Disability Research. 49*, 820-826.

Farrell, P. (2000). The impact of research on developments in inclusive education. *International Journal of Inclusive Education*, *4*, 453-162.

Ferguson, D, L. (2008). International trends in inclusive education: the continuing challenge to teach each one and everyone. *European Journal of Special Needs Education*, *23*, 109-120.

Fitzpatrick, S., & Bussey, K. (2011). The development of the Social Bullying Involvement Scales. *Aggressive Behavior*, *37*, 177-92.

Frostad, P., & Pijl, S. (2007). Does being friendly help in making friends? The relation between the social position and social skills of pupils with special needs in mainstream education. *European Journal of Special Needs Education*, *22*, 15-30.

Georgiou, S. N. (2008). Bullying and victimization at school: the role of mothers. *The British Journal of Educational Psychology*, *78*, 109-25.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*, 581-586.

Hatami, G., Motamed, N., & Ashrafzadeh, M. (2009). Confirmatory factor analysis of Persian adaptation of multidimensional students' life satisfaction scale (MSLSS). *Social Indicators Research*, *98*, 265-271.

Henson, R. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counselling, 34*, 177-189.

Hu, L, T., & Bentler, P.M., (1998). Fit indices in covariance structure modelling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

Humphrey, N. & Squires, G. (2011). *Achievement for All National Evaluation: Final Report*. Nottingham: Department for Education

Humphrey, N,, & Symes, W. (2010). Responses to bullying and use of social support among pupils with autism spectrum disorders (ASDs) in mainstream schools: a qualitative study. *Journal of Research in Special Educational Needs*, *10*, 82-90.

Kamphaus, R. (2004). *Behaviour assessment for children (2nd Ed)*. Circle Pines: AGS Publishing.

Kim, S., & Kamphaus, R. (2010). Is aggression the same for boys and girls? Measurement invariance with confirmatory factor analysis and item response theory. *School Psychology Quarterly*, *25*, 45-61.

Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2012). Validating the student-teacher relationship Scale: Testing factor structure and measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology*, *50*, 215-34.

Koster, M., Minnaert, A., Nakken, H., Pijl, S., & van Houten, E. J. (2011). Assessing social participation of students with special needs in inclusive education: Validation of the social participation questionnaire. *Journal of Psychoeducational Assessment*, *29*, 199-213.

Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *The British Journal of Educational Psychology*, *76*, 781-801.

Lee, T., & Cornell, D. (2009). Concurrent validity of the Olweus Bully/Victim Questionnaire. *Journal of School Violence*, *9*, 56-73.

Martlew, M., & Hodson, J. (1991). Children with mild learning difficulties in an integrated and in a special school: Comparisons of behaviour, teasing and teachers' attitudes. *British Journal of Educational Psychology*, *61*, 355-372.

Mishna, F. (2003). Learning disabilities and bullying: Double jeopardy. *Journal of Learning Disabilities*, *36*, 336-347.

Mooney, A., Owen, C., & Statham, J. (2008). *Disabled Children: Numbers, Characteristics and Local Service Provision*. London: Department for Children, Schools and Families.

Nabuzoka, D., & Smith, P. K. (1993). Sociometric status and social behaviour of children with and without learning difficulties. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *34*, 1435-48.

Porter, J., Daniels, H., Feller, A., & Georgeson, J. (2009). Collecting disability data from parents. *Research Papers in Education*, *i-first*, 1-17.

Riddick, B. (2012). Labelling learnings with "SEND": The good, the bad and the ugly. In D. Armstrong & G. Squires (Eds.), *Contemporary issues in Special Educational Needs* (pp. 25-34). Glasgow: McGraw Hill.

Rose, C. a., Monda-Amaya, L. E., & Espelage, D. L. (2010). Bullying Perpetration and Victimization in Special Education: A Review of the Literature. *Remedial and Special Education*, *32*, 114-130.

Ryser, G. R., Campbell, H. L., & Miller, B. K. (2010). Confirmatory factor analysis of the scales for diagnosing Attention Deficit Hyperactivity Disorder (SCALES). *Educational and Psychological Measurement*, *70*, 844-857.

Sebba, J., Peacock, A., DiFinizio, A., & Johnson, M. (2011). Personalisation and special educational needs. *Journal of Research in Special Educational Needs*, *11*, 203-224.

Smith, P, K., Cowie, H., Olafsson, R, F., and Liefooghe, A. (2002). Definitions of bullying: A comparison of terms used, and age and gender difference, in a fourteen- country international comparison. *Child Development, 73*, 1119-1133.

Solheim, E., Berg-Nielsen, T. S., & Wichstrom, L. (2011). The three dimensions of the student-teacher relationship scale: CFA validation in a preschool sample. *Journal of Psychoeducational Assessment*, *30*, 250-263.

Tabachnick, B., & Fidell, L. (2007). *Using Multivariate Statistics (5th ed.).* Boston: Pearson.

Tarshis, T. P., & Huffman, L. C. (2007). Psychometric properties of the peer interactions primary school (PIPS) questionnaire. *Journal of Developmental and Behavioral Pediatrics*, *28*, 125–32.

Taylor, L., Hume, I., & Welsh, N. (2010). Labelling and self-esteem: the impact of using specific vs. generic labels. *Educational Psychology*, *30*, 191-202.

Terwee, C. Bot, S. de Boer, M. van der Windt, D. Knol, D. Dekker, J. Bouter, L. et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42.

Tetler, S., & Baltzer, K. (2011). The climate of inclusive classrooms: The pupil perspective. *London Review of Education*, *9*, 333-344.

Van Eck, K., Finney, S. J., & Evans, S. W. (2010). Parent Report of ADHD Symptoms of Early Adolescents: A Confirmatory Factor Analysis of the Disruptive Behavior Disorders Scale. *Educational and Psychological Measurement*, *70*(6), 1042-1059.

Wiener, J., & Mak, M. (2009). Peer victimization in children with attention-deficit/hyperactivity disorder. *Psychology in the Schools*, *46*, 116-131.

Wigelsworth, M., Humphrey, N., Lendrum, A., & Kalambouka, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, *26*, 173-286.

Williams, T., Lamb, B., Norwich, B., & Peterson, L. (2009). Special Educational Needs has outlived its usefulness: a debate. *Journal of Research in Special Educational Needs*, *9*, 199-217.

Wright, P. (2004). *The Individuals with disabilities education improvement act of 2004: Overview explanation and comparison*. Retrieved from http://www.wrightslaw.com/idea/idea.2004.all.pdf

Van Eck, K., Finney, S. J., & Evans, S. W. (2010). Parent report of ADHD symptoms of early adolescents: A confirmatory factor analysis of the disruptive behavior disorders scale. *Educational and Psychological Measurement*, *70*, 1042–1059.

Van Roekel, E., Scholte, R. H. J., & Didden, R. (2010). Bullying among adolescents with autism spectrum disorders: prevalence and perception. *Journal of Autism and Developmental Disorders*, *40*, 63-73.

Appendix 1 the Wider Outcomes Survey for Teachers (WOST)

Behaviour
The pupil cheats and tells lies
The pupil takes things that do not belong to him/her
The pupil breaks or spoils things on purpose
The pupil gets angry and has tantrums
The pupil gets in fights with other children
The pupil says nasty things to other children

Bullying
The pupil is picked on by other children
The pupil is hurt by other children (e.g. gets pushed or kicked)
The pupil is called names or teased by other children
Other children spread unkind gossip about the pupil
Other children stop the pupil from joining in their games and activities at break times
The pupil is actively disliked by other children
Other children stop the pupil from joining in during class activities

Positive Relationships
The pupil can compromise with other children (e.g. take turns)
The pupil is helpful towards others
The pupil is popular with other children
The pupil can compromise with teachers (e.g. will complete a difficult task before moving on to a preferred activity)
The pupil is kind towards others
The pupil makes friends easily
The pupil can join in other children's activities

Table 1
Characteristics of Sample and Comparison with National Averages of pupils

identified with SEND (school action plus and statement)

	Sample	National Average (of SEND)*
% male	71.4	69
% EAL	16	12
%Eligible for free school meals	31.3	28.4
%Behavioural, Emotional, and Social Difficulties	20.6	22.7
%Cognition and Learning	45.4	41.1
% Communication and Interaction	27.0	24.4
%Sensory and/or Physical	7.0	7.3

*(figures obtained from the Department of Education, 2010)

Table 2. Descriptive Statistics and Reliability by SEND Classification

Classification of SEN		M	SD	α
Total sample (n = 6164)	Behaviour	0.62	0.73	.90
	Relationships	2.10	0.58	.92
	Bullying	0.53	0.62	.92
Behavioural, Emotional, and Social Difficulties (n = 1119)	Behaviour	1.22	0.84	.90
	Relationships	1.78	0.58	.91
	Bullying	0.81	0.69	.91
Cognition and Learning (n = 3675)	Behaviour	0.49	0.62	.89
	Relationships	2.20	0.55	.92
	Bullying	0.47	0.58	.92
Communication and Interaction (n = 1070)	Behaviour	0.54	0.64	.87
	Relationships	2.00	0.59	.91
	Bullying	0.50	0.59	.91
Sensory and/or Physical (n = 300)	Behaviour	0.40	0.61	.90
	Relationships	2.26	0.56	.92
	Bullying	0.36	0.50	.90

Table 3. Fit Indices for Models by SEND Classification

	x^2	df	CFI	TLI	AGFI	RMR	RMSEA
'Ideal' fit indices	-	-	>.9	>.9	>.9	≈0	≈0
Overall fit indices	13631.4	167	.858	.838	.734	0.04	.114
Behavioural, Emotional, and Social Difficulties	2791.1	167	.833	.810	.711	.066	.119
Cognition and Learning	7843.2	167	.860	.841	.741	.112	.033
Communication and Interaction	2394.3	167	.845	.823	.736	.042	.112
Sensory and/or Physical	957.4	167	.824	.799	.667	.042	.126

Table 4. Discriminative validity by SEND classification

Behaviour	Bullying	Positive relationships
BESD > ASD (*d*=0.55***) BESD > Other SEND (*d*=1.02***) ASD > Other SEND (*d*=0.47**)	BESD > ASD (*d*=0.26*) BESD > Other SEND (*d*=0.55**) ASD > Other SEND (*d*=0.28*)	BESD < Other SEND (*d*=0.72***) ASD < Other SEND (*d*=0.84***)

*indicates a 'small' effect size
** indicates a 'medium' effect size
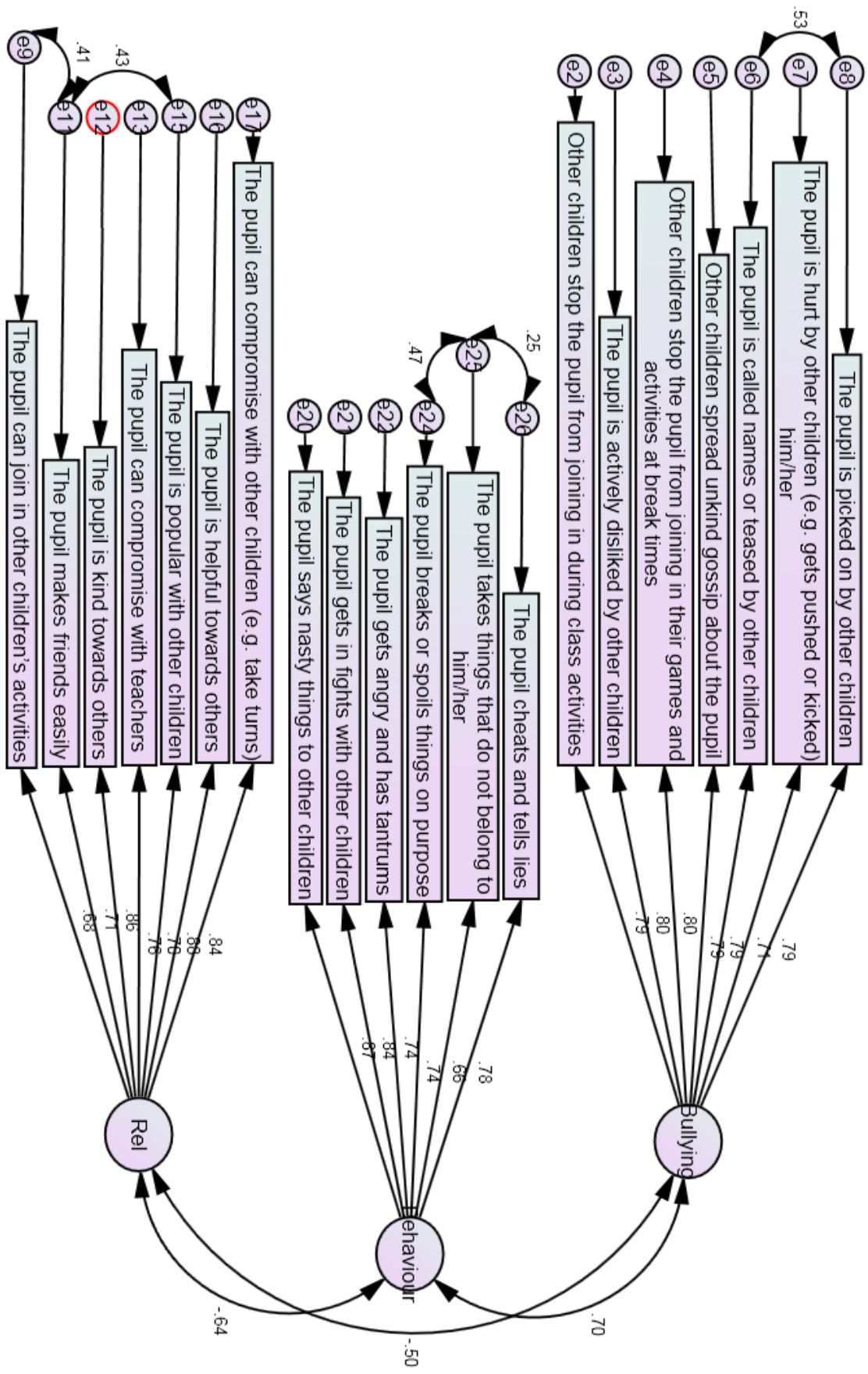*** indicates a 'large' effect size (Cohen, 1992)

Figure 1. Diagram of factor loadings (standardised estimates)