STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract p2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found p2
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported p3-5
Objectives	3	State specific objectives, including any prespecified hypotheses p5-6
Methods		
Study design	4	Present key elements of study design early in the paper p4, 6-10
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection p6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants p6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable p7-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group p7
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at p6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why p7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding p8-10
		(*b*) Describe any methods used to examine subgroups and interactions p8-10
		(*c*) Explain how missing data were addressed p8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy p8
		(*e*) Describe any sensitivity analyses p8
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed p7-8
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders p7-8
		(b) Indicate number of participants with missing data for each variable of interest p10
Outcome data	15*	Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included p10-15
		(b) Report category boundaries when continuous variables were categorized p7-8

		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses p10-15
Discussion		
Key results	18	Summarise key results with reference to study objectives p15-20
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias p19
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p15-19
Generalisability	21	Discuss the generalisability (external validity) of the study results p19
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p20

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Detecting mental health problems after paediatric Acquired Brain Injury: A pilot Rasch analysis of the Strengths and Difficulties Questionnaire

Robyn Henrietta McCarron[1,2], Fergus Gracey[1,3,4], and Andrew Bateman[2,5]

1. The Cambridge Centre for Paediatric Neuropsychological Rehabilitation, Cambridgeshire and Peterborough NHS Foundation Trust (CPFT)
2. Department of Psychiatry, University of Cambridge
3. National Institute of Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) East of England Programme
4. Department of Clinical Psychology, University of East Anglia
5. School of Health and Social Care, University of Essex

Author note

Robyn H. McCarron, Academic and Clinical Fellow in Psychiatry, Cambridgeshire and Peterborough NHS Foundation Trust and The Department of Psychiatry, University of Cambridge.

This research was supported by Addenbrookes' Charitable Trust under Grant 9980.

We have no known conflicts of interest to declare.

Correspondence address: The Cambridge Centre for Paediatric Neuropsychological Rehabilitation (CCPNR), Brookside Clinic, 18D Trumpington Road, Cambridge. CB2 8AH. Email: robyn.dudley@cantab.net

Word count

Abstract: 200

Main text: 5401

Abstract

The parent-reported Strengths and Difficulties Questionnaire (SDQ-P) is commonly used to assess for mental health problems, but its psychometric properties have not been studied in the paediatric Acquired Brain Injury (ABI) population. This study investigated the properties of the SDQ-P and its subscales in this population using Rasch analysis. One hundred and forty-three SDQ-Ps and 123 Impact Supplements were analysed. Sixty-nine percent of SDQ-Ps were completed by female carers, 59% of young people were male, and 58% had Traumatic Brain Injury (TBI). In this population the SDQ-P Total Difficulties Scale and the Conduct Problems subscale showed questionable construct validity. The individual subscales and Impact Supplement did not meet the criteria for reliability. Disordered thresholds were seen for two items. The individual subscales showed mistargeting and 13-24% person misfit. Two items were significantly underdiscriminating. There was differential item functioning with age and time post-injury, and local dependence between subscale items. The Total Difficulties scale was multidimensional. The most easily endorsed items were in keeping with common symptoms of brain injury. These findings suggest the SDQ-P in its current form may not be a reliable and valid assessment measure for mental health difficulties in the paediatric ABI population and requires further investigation.

Keywords

Acquired Brain Injury, Rasch analysis, construct validity, mental health, psychosocial

Main Text

Introduction

The development of mental health problems following paediatric Acquired Brain Injury (ABI) is a significant public health problem. ABI affects an estimated 40,000 children in the UK each year (National Health Service (NHS) England, 2013), and a clear association has been shown between ABI and the development of new-onset psychiatric disorders (Max et al., 2012; Schwartz et al., 2003), with 61% (Max et al., 1997) of children and adolescent developing psychiatric disorders in the two years post injury. These new onset psychiatric disorders cover the spectrum of psychiatric diagnoses and include depressive disorders (Luis & Mittenberg, 2002; Max et al. 2012), anxiety disorders including post-traumatic stress disorder (PTSD) (Hajek et al. 2010; Max et al. 2015), secondary attention-deficit hyperactivity disorder (S-ADHD) (Levin et al. 2007; Sinopoli, Schachar, & Dennis, 2011), and behavioural disorders (Schwartz et al., 2003).

The importance of early recognition and treatment of mental health problems in improving a young person's prognosis is widely recognised (Honeyman, 2007; Membride, 2016). However, in a study of young people with Traumatic Brain Injury (TBI) Heubner et al. (2018) found that 68% had unmet mental health needs. Poor consensus on how psychiatric difficulties after ABI are defined (McKinlay, Grace, Horwood, Fergusson, & MacFarlane, 2009; Ornstein et al., 2013; Soo, Tate, & Brookes, 2014), a lack of understanding of the impact of brain injuries on young people (Kirk, Fallon, Fraser, Robinson, & Vassallo, 2015), and a poverty of standardised assessment measures specifically designed or validated for use in the paediatric ABI population (McCarron, 2017; Soo et al. 2014) are likely barriers to the timely recognition of mental health problems in young people with ABI.

The parent-reported Strengths and Difficulties Questionnaire (SDQ-P) (Goodman, 1997; Stone, Otten, Engels, Vermulst, & Janssens, 2010) is commonly used by paediatric ABI services and in ABI research (Brooks et al., 2019; Pastore et al., 2018) to screen for mental health and psychosocial problems in young people. The SDQ-P is a brief behavioural screening questionnaire that assesses a range of positive and negative psychological attributes in young people aged 4-17 years old and has been demonstrated to work as a dimensional measure of mental health (Goodman & Goodman, 2012). Its validity in the general population (Kersten et al., 2016; Stone et al., 2010), relative ease of use, brevity, multiple language translations and free availability make it an attractive assessment measure for services. However, to our knowledge the psychometric properties of the SDQ-P in the paediatric ABI population have not yet been explored. The crossover between some common brain injury symptoms, such as headaches (Babcock et al., 2013), impulsivity (Wassenberg, Max, Kindgren, & Schatz, 2004), impaired attention (Catroppa, Anderson, Morse, Haritou, & Rosenfeld, 2006; Yeates et al, 2005) and social difficulties (Ewing-Cobbs, Prasad, Mendez, Barnes, & Swank, 2013; Yeates et al. 2007) and symptoms of mental health difficulties assessed by the SDQ-P, may negatively affect its validity and reliability in the paediatric ABI population. Furthermore, the heterogeneity (McCarron, 2017) of the paediatric ABI population may impair the accuracy with which results can be interpreted across a range of ages, injury types and times post-injury.

Traditionally, the psychometric properties of assessment measures in terms of validity and reliability have been investigated using methods based on Classical Test Theory (CTT). The Rasch Measurement Model (Rasch, 1960) is a modern psychometric technique that falls within the parameters of Item Response Theory (IRT) (Hambleton, Swaminathan, & Rogers, 1991). Unlike CTT the Rasch model has the advantage of not assuming the equivalence between ordinal and interval scales (Hobart & Cano, 2009). Nor does it rely on the assumption that the observed scores are composed of the true score and an error (neither of which can be determined), in order to estimate the reliability of the observed score. Instead, the Rasch model is based on assumptions that readily make sense within a real-world context. It tests the assumptions that people respond in a probabilistic but ordered manner based on both their underlying traits (be it ability or disease severity) and the level or difficulty being assessed by an item/question. It maintains that an assessment measure should not be biased towards individuals with certain characteristics or previous responses, and it argues that for a total score to be meaningful it needs to be reflective of a single unidimensional construct. Rasch analysis has been demonstrated to be an insightful method for examining the psychometric properties of rating scales in different populations, including in people with ABI (Bateman, Teasdale, & Willmes, 2009; Simblett, Gracey, Ring, & Bateman, 2015).

Whilst studies utilising CTT (Kersten et al., 2016; Stone et al., 2010) have supported the validity of the SDQ in the general population, it has been found to have poor validity when investigated using Rasch analysis in New Zealand pre-schoolers (Kersten et al., 2018). Combined with the potential population-specific issues this creates a clear rationale for undertaking a Rasch analysis of the psychometric properties of the SDQ-P in the paediatric ABI population. Given the current balance of evidence in support of the SDQ-P, undertaking a pilot study to assess the case for a larger-scale study of its construct validity and reliability in this population was warranted.

Aims

This study aimed to:

1. Undertake a pilot Rasch analysis of the SDQ-P in the paediatric ABI population to identify if a larger-scale study of its construct validity and reliability in this population is required.

2. To qualitatively interpret the findings of the Rasch analysis in the context of what is known about the psychosocial consequences of paediatric ABI.

The assumptions to be tested by the Rasch analysis are:

1) People respond probabilistically, so that parents are most likely to endorse commonly reported symptom, and parents of children with greater mental health problems will endorse more items relating to psychosocial difficulties.

2) Parents will respond to categories in an ordered manner reflecting symptom severity.

3) There should be no differential item functioning, meaning that the probability of a parent endorsing a symptom should be independent of their or their child's demographic or injury related factors.

4) Items should be independent, so that endorsing one symptom should not increase the probability of a parent endorsing another symptom.

5) The SDQ-P subscales and total difficulties scale should be unidimensional, indicating that they measure a single construct.

6) The SDQ-P should be targeted to capture the range of the degree of psychological strengths and difficulties reported by parents of young people with ABI.

Methods

Ethical approval

This study was approved by the local National Health Service (NHS) Trust's quality improvement department. All data was obtained and analysed in compliance with data protection legislation and the Trust's regulations.

Sample

The sample was young people with ABI obtained through consecutive referrals to a specialist community paediatric neuropsychological rehabilitation service based within the UK NHS, between 2010 and 2019. This time period, and the resulting sample size, was chosen to permit a review of data and clinical practice moving in to the second decade of the service. The service accepts young people up to the age of 19 with any type or severity of ABI, whose needs cannot be met locally by non-specialist community services.

It is routine clinical practice that an SDQ-P is completed as part of the initial assessment of young people attending the service. One hundred and forty-three out of 167 (86%) young people attending the service had a completed SDQ-P. Of the 143 respondents 123 (86%) had completed the Impact Supplement. Missing SDQ-Ps and Impact Supplements were due to non-return/non-completion of questionnaires, with no systematic differences between young people with and without completed questionnaires. These samples were representative of the young people attending the service, including a range of ages, injury types, and time post-injury (TPI). The small sample size is reflected in the pilot nature of this analysis.

Fifty-nine percent of young people were male. The age at assessment ranged from 4-17 years (median 13 years), covering the full age range of the SDQ-P. Injury type was 58% TBI, 15% infection, 11% vascular, 7% tumour, 6% autoimmune and 3% other causes. TPI ranged from 0-17 years (median 6 years). The SDQ-P had been completed by a female carer in 69% of cases, a male carer in 13%, and carer sex was unknown in 19% of cases.

Measure

The 25 item SDQ-P and the Impact Supplement were completed by participants as part of routine clinical practice. The SDQ-P is a polytomous scale in which items are endorsed as “Not True”, “Somewhat True” or “Certainly True” and are scored 0, 1 and 2 respectively, apart from five items which are reversed scored. The SDQ-P consists of four subscales of five items each, assessing Emotional Symptoms, Conduct Problems, Hyperactivity and Peer Problems, that are combined into a Total Difficulty Scale measuring negative psychological attributes, the total score of which can be interpreted as an indication of mental health problems. The Prosocial Scale measures positive psychological attributes. The additional Impact Supplement consists of five polytomous items assessing overall distress and social impairment. These items are endorsed as “Not at all”, “Only a little”, “Quite a lot” or “A great deal”, scoring 0, 0, 1 and 2 respectively.

Data extraction and processing

Assessment data is routinely collated by the service on a Microsoft Excel spreadsheet. Data on SDQ-P scores, parent sex, child sex, type of injury, age at injury and time-post injury (TPI) was extracted for analysis from this spreadsheet. Where data was missing clinical records were reviewed to try and identify this information.

Demographic and injury data was categorised to allow person factors to be investigated. Type of injury was classified as TBI or non-TBI. Child age categories were defined as preschool or primary school (<11 years), secondary school (11-<16 years) and post-compulsory education (≥16 years). Time-post injury was categorised as less than 2 years, 2-5 years and 5 years and over, in keeping with commonly used research categories (LeBlanc et al. 2008, Ponsford et al. 2014).

Data was prepared for Rasch analysis using Microsoft Excel.

Missing data

Responses to the SDQ-P that contained missing item or person factor data were retained in the analysis. Missing data was recorded using a * symbol. Missing data analysis was not performed due to the low rates of missing data at an item level.

Data analysis

Rasch analysis was performed using RUMM2030 software (Andrich, Sheridan, & Luo. 2009) and following reporting guidelines proposed by Smith, Linacre and Smith (2003) and Tennant and Conaghan (2007). Rasch analysis was performed separately on each subscale and on the Total Difficulties scale.

The choice of Rasch model was determined using the likelihood ratio test, with the Masters Partial Credit Model (Masters, 1982) (Figure 1.) being used where the chi-square statistic was significant, and the Andrich Rating Scale Model (Andrich, 1978) where it was not (Figure 1.).

[INSERT FIGURE 1. HERE]

Reliability was primarily assessed using the Person Separation Index (PSI). This is a measure of the replicability of person ordering if the same sample were assessed on a corresponding set of items measuring the same underlying construct. Values above 0.7 and 0.85 respectively indicate sufficient reliability for group and individual use (Tennant & Conaghan, 2007). Cronbach's alpha was reported for completeness, although it has been found to be a poor indicator of reliability in Rasch analysis (Smith et al., 2003) and the SDQ (Kersten, 2018).

Overall fit to the Rasch model was evaluated using the total chi-squared item-trait interaction statistic. This examines the difference between the observed and expected values across different groups representing different degrees of psychosocial symptom severity. A non-significant result ($p \geq 0.05$) indicates that the items are assessing the same underlying construct, so fit to the Rasch model. Category structure was examined for threshold ordering. Disordered thresholds were addressed by collapsing categories and re-scoring items.

Sources of misfit were then further investigated. Individual person fit was investigated, with people considered misfitting if their fit residuals fell outside the range of -2.5 to +2.5 and/or their chi-squared statistic was significant ($p < 0.05$ corrected for multiple comparisons using the Bonferroni correction). A high negative fit residual indicates that people are responding in a highly predictable manner (e.g. consistently endorsing the highest or lowest scores for an item), and a high positive fit residual indicates highly unpredictable responding (e.g. a random pattern of responding) (Wright, 1995). Due to the small sample size, and pilot nature of this analysis, misfitting people were retained in the analysis.

Individual item fit was then examined as per person fit. Fit residuals below -2.5 indicate than an item is over-discriminating between levels of psychological symptoms, whilst residuals above +2.5 indicate the item is under-discriminating. Misfitting items were then removed from the analysis.

Items were then assessed for differential item functioning (DIF), or item bias. It was examined if parents responded differently to items on the SDQ-P based on parent sex, child sex, type of injury, age at injury or TPI. Uniform DIF is present if there is a systematic difference in responding across the whole range of the symptom being measured, and non-uniform DIF is where differences vary across the range of symptom severity. Differential item functioning was determined using a two-way analysis of variance (ANOVA) calculation, with significant DIF being determined by a p value < 0.05 adjusted by the Bonforroni correction. The impact of removing items displaying significant DIF was explored.

Local item dependency was examined using a residual correlation matrix to assess whether the probabilities of endorsing separate items were linked. Correlations greater than 0.3 above the average between pairs of items indicated dependency. The effect of combining locally dependent items into a subtest was explored.

Unidimensionality was assessed following these changes using the method proposed by Smith (2002), to determine if a single underlying construct was being measured. Principal component analysis was performed to separate the items into two groups, from which t-tests of the person estimates were performed. If a significant difference ($p < 0.05$) was seen in under 5% of the sample unidimensionality was assumed. A binomial test was performed and if the lower 95% confidence interval (CI) for the proportion of significant t-tests was less than 5% then unidimensionality was further supported.

The final targeting of the scale was assessed to determine whether the full spectrum of disease and symptom severity could be adequately captured, or whether it failed to distinguish between individuals at the upper (ceiling effect) and lower (floor effect) levels. This was done through an assessment of mean person location and visual inspection of the person-item threshold distribution plot.

Results

Missing data

Of the 143 SDQ-Ps, 128 (90%) were complete for the five main subscales. There was no missing data in the 123 completed Impact Supplements. For individual items the rates of missing data was low, ranging from 0% to 2%, with item 25 “Sees tasks through to the end, good attention span” having the highest rate of missing data, suggesting a systematic pattern of missing data due to this being the final item on the questionnaire.

Choice of model

Based on significant Likelihood Ratio Tests the Masters Partial Credit Model was used for all scales except the Hyperactivity subscale ($\chi 2(3) = 5.73$, p = 0.13) and the Impact Supplement ($\chi 2(3) = 3.67$, p = 0.30) where the Andrich Rating Scale Model was used.

Emotional Symptoms subscale

The Emotional Symptoms subscale showed overall fit to the Rasch model and ordered thresholds, but insufficient reliability (Table 1.). Seventeen percent of respondents showed misfit, with 83% of these being located extreme person locations (of these 80% were at an extreme negative location), and the remainder having high negative fit residuals indicating highly predictable responding. Item 8 “Many worries, often seems worried” was the easiest item to endorse and item 13 “Often unhappy, downhearted or tearful” was the hardest. Item 3 “Often complains of headaches, stomach-aches…” showed significant misfit with a fit residual of 3.02, indicating that this item is under-discriminating. This item was removed from the analysis, further improving the item-trait interaction fit. There was no DIF or local dependency. The modified subscale (Table 1.) showed unidimensionality, but there was some mistargeting with a mean person location of -0.23 (*SD* = 1.75) and both a floor and ceiling effect on visual inspection (Figure 2a.).

[INSERT TABLE 1. HERE]

[INSERT FIGURE 2. HERE]

Conduct Problems subscale

The Conduct Problems subscale showed poor reliability and poor fit to the Rasch model (Table 1.). Item 22 “Steals from home, school or elsewhere” showed disordered thresholds (Figure 3a.), so this item was rescored combining responses 1 and 2.

[INSERT FIGURE 3 HERE]

Fifteen percent of people showed misfit due to extreme location, with 90% of these being located at an extreme negative location. Item 5 “Often has temper tantrums or hot tempers” was the easiest to endorse and item 22 “Steals from home, school or elsewhere” was the hardest. There were no misfitting items or DIF. There was local dependency between items 18 “Often lies or cheats” and 22. Combining these items into a subscale was explored, but this did not improve fit and the items were felt to be qualitatively different, so the individual items were retained. The modified subscale (Table 1.) was unidimensional, but the mean person location showed clear mistargeting (mean = -1.043, *SD* = 1.681) with both a floor and ceiling effect on visual inspection (Figure 2b.).

Hyperactivity subscale

The Hyperactivity subscale showed overall fit to the Rasch model and ordered thresholds but did not meet the criteria for sufficient reliability (Table 1.). Thirteen percent of respondents were misfitting due to extreme locations, of which 95% were located at high positive locations. There were no misfitting items. Item 15 “Easily distracted, concentration wanders” was the easiest to endorse and item 10 “Constantly fidgeting or squirming was the hardest”. There was no DIF. Local dependency was seen between item 2 “Restless, overactive, cannot stay still for long” and 10, between item 15 and 25 “Sees tasks through to the end, good attention span” and between items 21 “Thinks things out before acting” and item 25. The dependent items were combined into subscales but this did not improve the fit to the Rasch model and resulted in extremely disordered thresholds, so the individual items were retained. The subscale showed unidimensionality (Table 1.). A floor and ceiling effect (Figure 2c.) was seen with a mean person location of 0.69 (*SD* = 1.46) supporting mistargeting.

Peer Problems subscale

The Peer Problems subscale showed overall fit to the Rasch model and ordered thresholds but poor reliability (Table 1.). Seventeen percent of people were misfitting due to extreme locations, with 96% of these having extreme negative locations. Item 23 “Gets on better with adults than children” was easiest to endorse, and item 11 “Has at least one good friend” (reverse scored so higher scores indicate friendship difficulties) was hardest to endorse, with no misfitting items. Significant non-uniform DIF by age was seen for items 11 ($F(4) = 4.25, p = 0.0031$) (Figure 4a.) and 19 “Picked on or bullied by other children” ($F(4) = 6.32$, $p = 0.00013$) (Figure 4b). Removing these items did result in some improvement in the chi-squared interaction, but led to a major reduction in the PSI, so the items were retained.

[INSERT FIGURE 4. HERE]

There was no response dependency and the subscale satisfied the criteria for unidimensionality (Table 1.). The subscale showed mistargeting with a mean person location of -0.95 (*SD* = 1.46) and a floor and ceiling effect (Figure 2d.).

Prosocial subscale

The Prosocial subscale approached the acceptable criteria for reliability, showed overall fit to the Rasch model and all thresholds were ordered (Table 1). Sixteen percent of people were misfitting due to extreme values, with 91% of these having an extreme positive location. There were no misfitting items, with item 17 “kind to younger children” being the easiest and item 20 “Often volunteers to help others” being the hardest to endorse. There was no DIF or local dependency, and the criteria for unidimensionality were achieved (Table 1). The subscale was mistargeted with a mean person location of 1.18 (*SD* = 1.8) and a floor and ceiling effect (Figure 2e.).

Total Difficulties scale

The Total Difficulties scale showed good reliability but fit to the Rasch model was not supported (Table 1.). Disordered thresholds were seen for two items from the conduct subscale, item 12 (Figure 3b.) “Often fights with other children or bullies them” and item 22 “Steals from home, school or elsewhere” (Figure 3c.) with a score of 1 never being most likely. These items were rescored combining responses 1 and 2.

There were three misfitting people (2% of the sample), due to one high positive and one high negative fit residual and pone extreme negative location. Item 15 was the easiest to endorse and item 22 the hardest. Item 3 “Often complains of headaches, stomach-aches…” (Emotional Symptoms) showed significant with a high positive fit residual (2.60) and significant chi-square ($\chi 2(2) = 19.56$, $p < 0.0005$), and item 23 “Gets on better with adults than with other children” (Peer Problems) also had a high positive fit residual (2.53). These two items were removed. Uniform DIF by TPI was seen for item 19 “Picked on or bullied by other children” ($F(2) = 7.80$, $p = 0.00063$), with increased bullying being reported with increasing TPI (Figure 4c). Removing this item did not improve fit to the Rasch model so it was retained.

Significant local dependency was seen within each of the subscales composing the Total Difficulties scale, with local dependency affecting all items on the Emotional Symptoms and Hyperactivity subscales, and items 18 “Often lies or cheats” and 22 “Steals from home, school or elsewhere” and items 11 “Has at least one good friend” and 14 “Generally liked by other children” on the Conduct and Peer Problems subscales. There was no local dependency on items across subscales. Various options of combining locally dependent items into subscales were explored, but in no iteration could fit to the Rasch model be improved whilst preserving the reliability of the scale, so the individual items were retained. The final scale did not meet the criteria for unidimensionality (Table 1.) but showed reasonable targeting with a mean person location of 0.326 (*SD* = 1.172) (Figure 2f.).

Impact Supplement

The Impact Supplement showed fit to the Rasch model and ordered thresholds but did not reach the criteria for reliability (Table 1.). Twenty-four percent of respondents showed misfit to the sample. Of these 60% had an extreme negative location, 20% had an extreme positive location and 20% had high negative fit residuals. Item 4 "Difficulties interfere with classroom learning" was the easiest to endorse, and item 5 "Difficulties interfere with leisure activities" was the hardest to endorse, with no misfitting items. Uniform DIF by TPI was seen for item 3 "Difficulties interfere with friendships" (Figure 4d.) with difficulties having the greatest impact on friendships in the group over 5 years post-injury and the least impact in the group 2-5 years post-injury. However, removing this item resulted in poorer fit and reliability so it was retained.

There was no local dependency between items, and the supplement fulfilled the criteria for unidimensionality (Table 1.). The scale showed a floor and ceiling effect (mean person location = -0.296, *SD* = 1.59) (Figure 2g.).

Discussion

In the paediatric ABI population studied in this analysis, the SDQ-P showed questionable construct validity with a significant chi-squared item-trait interaction for the Total Difficulties scale and the Conduct Problems subscale. This has clinical implications, as the SDQ-P Total Difficulties scale may not be a valid screening measure for detecting mental health problems in this population. Whilst the Total Difficulties scale showed sufficient reliability to distinguish between individuals, the SDQ-P subscales and Impact Supplement did not meet the criteria for reliability. This is likely an effect of the small number of items on each subscale, but clinicians should be mindful of this when interpreting results. The subscales and Impact Supplement showed unidimensionality, but the Total Difficulties scale appeared multidimensional. This is unsurprising as the scale is assessing difficulties across the different traits of Emotionality, Conduct Problems, Hyperactivity and Peer Problems. However, clinically and in research this is problematic as a young person with an ABI scoring highly based on Conduct and Peer Problems cannot be considered equivalent to one with the same total score resulting from Hyperactivity and Emotional Symptoms.

The individual subscales showed 13-24% person misfit. This reflects the diversity of post-ABI presentations and symptoms, with subsets of individuals reporting very low or very high degrees of difficulties in some areas. For example, the 90% of misfitting people on the Conduct problems subscale had extreme negative locations (low degree of difficulty) whereas 95% of misfitting people on the Hyperactivity scale had extreme positive locations (high degree of difficulty).

The person misfit is reflected in the targeting of the subscales, which all showed a floor and ceiling effect. Clinically, this means that the Total Difficulties score may not adequately detect individuals with mental health problems in the paediatric ABI population, and individuals scoring highly on subtests require further in-depth assessment in that area. The negative mean person locations on the Peer and Conduct Problems subscales suggests that these subscales may be insufficiently sensitive to detect individuals with problems in these areas in this population and may require modification with the addition of items assessing lower-level difficulties. Conversely, the prevalence of problems in attention and concentration following paediatric ABI (Catroppa, 2006; Yeates et al, 2005) may be contributing to the items on the hyperactivity scale being too easy, and more challenging questions, perhaps focussed on functioning or with additional items relating to cognition, are required to improve its specificity.

Two items were removed from the Total Difficulties scale due to misfit. Item 3 "Often complains of headaches, stomach-aches…" (Emotional Symptoms) item 23 "Gets on better with adults than with other children" (Peer Problems) are both underdiscriminating in the paediatric ABI population. Although somatic symptoms are commonly comorbid with mental health problems in children and adolescents (Kristensen, Oerbeck, Torgersen, Hansen, & Wyller, 2014; Masi, favilla, Millepiedi, & Mucci, 2000), the high prevalence of somatic symptoms (Yeats et al. 2003), especially headaches (Babcock et al. 2013), following paediatric ABI means that item 3 is unable to discriminate between young people with and without emotional symptoms following ABI. Problems with social adjustment, which is "The degree to which children get along with their peers" (Crick & Dodge, 1994, p.82), are common after paediatric ABI (Anderson et al., 2017), and the increased time spent with adults following ABI due to hospitalisation, rehabilitation and care needs may all contribute to young people with ABI getting on better with adults than children. This may contribute to item 23 being both underdiscriminating and the easiest item to endorse in its subscale. The easiest items to endorse across the other subscales are all in keeping with common post-brain injury symptoms in young people, such as anxiety (Max et al., 2015), attentional difficulties (Yeates et al, 2005) and emotional dysregulation (Finnanger et al., 2015). The impact of ABI on a young person's ability to participate in education (Andersson, Bellon, & Walker, 2016, Ylvisaker et al., 2001) is reflected by "Difficulties in classroom learning" being the easiest item to endorse on the Impact Supplement.

Non-uniform DIF by age was seen for two items when the Peer Problems subscale was individually examined, but these effects were no longer significant in the more reliable Total Difficulties scale analysis. However, a uniform pattern of DIF by TPI emerged for item 19 "Picked on or bullied by other children" with reported bullying increasing with increasing TPI. Young people with ABI are more at risk of being bullied or victimised (Hung et al., 2017), and with increasing TPI young people may be able to participate more fully in school and community life, which may result in their differences and difficulties becoming a more apparent target to their peers. Similarly, uniform DIF by TPI was also seen for item 3 “Difficulties interfering with friendships” on the impact supplement, with the greatest impact seen in the group over 5 years post-injury and the least impact in the group 2-5 years post injury. Friendship difficulties are common post-ABI (Prigatano & Gupta, 2006; Yeates et al., 2013), and an increased need for support with friendships in young people over 5 years post-injury has previously been identified in this setting (McCarron, Watson, & Gracey, 2019). A possible explanation for this would be that in the first 2 years post injury a young person’s post-acute physical and cognitive difficulties impact on their ability to participate and maintain friendships, but things might then improve through a process of rehabilitation, where available. However, after 5 years post-injury the young person is participating more alongside their peers and undergoing a period or psychological adjustment to reconcile their pre and post-injury senses of self (Gracey, Evans, & Malley, 2009). Through this they may become more aware of the differences between themselves and their peers, impacting further on friendships.

A high degree of local dependence was seen in the SDQ-P. Although items on the Emotional Symptoms and Peer Problems subscales appeared independent initially, this may have been a consequence of the small number of items on these scales as clear local dependence was seen when items were combined in the Total Difficulties scale. Local dependence of items assessing the same underlying traits may be expected, for example cognitive symptoms of inattention are likely to be associated with physical symptoms of hyperactivity and impulsivity, but this linking of items affects the Rasch analysis and any interpretation of a Total Difficulties score. There is clear qualitative similarity between many of the subscale items, such as between item 8 “Many worries, often seems worried” and item

24 “Many fears, easily scared” on the Emotional Symptoms subscale, and item 15 “ Easily distracted, concentration wanders” and item 25 “Sees tasks through to the end, good attention span” on the Hyperactivity subscale. Although combining items into subtests did not improve the properties of the SDQ-P in this analysis, it should be investigated further in a larger study, as creating a shorter form of the questionnaire for use in the paediatric ABI population would be of clinical utility and reduce the questionnaire burden on families undergoing often lengthy assessment processes. In addition, revision of the SDQ-P for the paediatric ABI population should also consider the viability of standard psychiatric classifications in this group and seek to identify item groupings or factors that match with common symptom clusters arising due to neuroanatomical disruption or psychological adjustment and adaptation.

Limitations

This study is limited by its small sample size. This is thought to have particularly impacted on the reliability of the results, especially when items were removed or combined into subscales. However, as a pilot analysis this study is felt to be of sufficient size to justify a large-scale study of the SDQ-P in the paediatric ABI population. Respondent bias may have been seen in the individuals completing the impact subscale. The study was conducted within a specialist service for young people with highly complex neuropsychological needs that cannot be met by non-specialist teams, so the results may not be generalisable to all young people with ABI in the community. The reader should be mindful that this study investigates the parent-reported SDQ so findings cannot be generalised to the self-report or teacher-report versions.

Conclusions

This pilot Rasch analysis of the SDQ-P, including its subscales and Impact Supplement, raises concerns about its construct validity and reliability in the paediatric ABI population. Issues were identified in overall fit to the Rasch model, disordered thresholds, misfitting people and items, DIF, local dependency, dimensionality and targeting. This requires further investigation in a larger study, particularly with regards to the impact of removing misfitting or differentially functioning items and combining locally dependent item into subtests. The findings of this study can be understood qualitatively in the context of what is known about the psychosocial consequences of paediatric ABI. Acquired Brain Injury in young people results in a diverse range of cognitive, psychological, social and physical symptoms, the complexity of which may evade capture by a standardised assessment measure and may need to be interpreted qualitatively and holistically within the context of a specialist interdisciplinary service.

Acknowledgements

We thank the young people and their families who have provided this data, and the clinical team who have collected it.

This work was supported by Addenbrookes' Charitable Trust under Grant 9980.

Disclosure of interest

There are no competing interests to declare.

References

Anderson, V., Beauchamp, M. H., Yeates, K. O., Crossley, L., Ryan, N., Hearps, S. J., & Catroppa, C. (2017). Social competence at two years after childhood traumatic brain injury. *Journal of Neurotrauma*, *34*(14), 2261-2271.

Andersson, K., Bellon, M., & Walker, R. (2016). Parents’ experiences of their child’s return to school following acquired brain injury (ABI): A systematic review of qualitative studies. *Brain Injury, 30*(7), 829-838.

Andrich, D. (1978). Rating formulation for ordered response categories *Psychometrika*, *43*, 561-573.

Andrich, D., Sheridan, B. S., & Luo, G. (2009). *RUMM2030: Rasch unidimensional models for measurement.* Perth, Western Australia: RUMM Laboratory.

Babcock L., Byczkowski T., Wade S.L., Ho M., Mookerjee S., & Bazarian J.J. (2013). Predicting postconcussion syndrome after mild traumatic brain injury in children and adolescents who present to the emergency department. *JAMA Pediatr*. *167*, 156–161

Bateman, A., Teasdale, T. W., & Willmes, K. (2009). Assessing construct validity of the self-rating version of the European Brain Injury Questionnaire (EBIQ) using Rasch analysis. *Neuropsychological Rehabilitation*, *19*(6), 941-954.

Brooks, B. L., Plourde, V., Beauchamp, M. H., Tang, K., Yeates, K. O., Keightley, M., Anderson, P., Désiré, N., Barrowman, N., Zemek, R., & PERC Concussion Team. (2019). Predicting psychological distress after pediatric concussion. *Journal of Neurotrauma, 36*(5), 679-685.

Catroppa, C., Anderson, V. A., Morse, S. A., Haritou, F., & Rosenfeld, J. V. (2006). Children’s attentional skills 5 years post-TBI. *Journal of Pediatric Psychology, 32*(3), 354-369.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74.

Ewing-Cobbs, L., Prasad, M. R., Mendez, D., Barnes, M. A., & Swank, P. (2013). Social interaction in young children with inflicted and accidental traumatic brain injury: relations with family resources and social outcomes. *Journal of the International Neuropsychological Society*, *19*(5), 497-507.

Finnanger, T. G., Olsen, A., Skandsen, T., Lydersen, S., Vik, A., Evensen, K. A. I., Catroppa, C., Håberg, A. K., Andersson, S., & Indredavik, M. S. (2015). Life after adolescent and adult moderate and severe traumatic brain injury: self-reported executive, emotional, and behavioural function 2–5 years after injury. *Behavioural Neurology, 2015*.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581-586.

Goodman, A., & Goodman, R. (2012). Strengths and Difficulties Questionnaire scores and mental health in looked after children. *The British Journal of Psychiatry*, *200*(5), 426-427.

Gracey, F., Evans, J. J., & Malley, D. (2009). Capturing process and outcome in complex rehabilitation interventions: A “Y-shaped” model. *Neuropsychological Rehabilitation*, *19*(6), 867-890.

Hajek, C.A., Yeates, K.O., Gerry Taylor, H., Bangert, B., Dietrich, A., Nuss, K.E., Rusin, J., & Wright, M. (2010). Relationships among post-concussive symptoms and symptoms of PTSD in children following mild traumatic brain injury. *Brain Injury*, *24*(2), 100-109.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA: Sage.

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment*, *13*(12):1-200

Honeyman, C. (2007). Recognising mental health problems in children and young people. *Paediatric Nursing*, *19*(8), 38–44.

Huebner, A. R. S., Cassedy, A., Brown, T. M., Taylor, H. G., Stancin, T., Kirkwood, M. W., & Shari, L. W. (2018). Use of Mental Health Services by Adolescents After Traumatic Brain Injury: A Secondary Analysis of a Randomized Controlled Trial. *PM&R*, *10*(5), 462-471.

Hung, A. H., Cassedy, A., Schultz, H. M., Yeates, K. O., Taylor, H. G., Stancin, T., Walz, N. C. & Wade, S. L. (2017). Predictors of long-term victimization after early pediatric traumatic brain injury. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *38*(1), 49.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, *40*(1), 64-75.

Kirk, S., Fallon, D., Fraser, C., Robinson, G., & Vassallo, G. (2015). Supporting parents following childhood traumatic brain injury: a qualitative study to examine information and emotional support needs across key care transitions. *Child: Care, Health and Development, 41*(2), 303-313.

Kristensen, H., Oerbeck, B., Torgersen, H. S., Hansen, B. H., & Wyller, V. B. (2014). Somatic symptoms in children with anxiety disorders: an exploratory cross-sectional study of the relationship between subjective and objective measures. *European Child & Adolescent Psychiatry*, *23*(9), 795-803.

LeBlanc, J., Feyz, M., Meyer, K., Duplantie, J., Thomas, H., Abouassaly, M., ... & Robinson, C. (2008). Long-term outcome after severe traumatic brain injury: the McGill interdisciplinary prospective study. T*he Journal of Head Trauma Rehabilitation*, *23*(5), 294-303.

Levin, H., Hanten, G., Max, J., Li, X., Swank, P., Ewing-Cobbs, L., Dennis, M., Menefee, D. S., & Schachar, R. (2007). Symptoms of attention-deficit/hyperactivity disorder following traumatic brain injury in children. *Journal of Developmental & Behavioral Pediatrics*, *28*(2), 108-118.

Luis, C. A., & Mittenberg, W. (2002). Mood and anxiety disorders following pediatric traumatic brain injury: a prospective study. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 270-279.

Masi, G., Favilla, L., Millepiedi, S., & Mucci, M. (2000). Somatic symptoms in children and adolescents referred for emotional and behavioral disorders. *Psychiatry*, *63*(2), 140-149.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

Max, J. E., Lopez, A., Wilde, E. A., Bigler, E.D., Schachar, R.J., Saunders, A., Ewing-Cobbs, L., Chapman, S.B., Yang, T.T., & Levin, H.S. (2015). Anxiety disorders in children and adolescents in the second six months after traumatic brain injury. *Journal of Pediatric Rehabilitation Medicine*, *8*(4), 345-355.

Max, J. E., Robin, D. A., Lindgren, S. D., Smith Jr, W. L., Sato, Y., Mattheis, P. J., Stierwalt, J., & Castillo, C. S. (1997). Traumatic brain injury in children and adolescents: psychiatric disorders at two years. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(9), 1278-1285.

Max, J. E., Wilde, E. A., Bigler, E. D., MacLeod, M., Vasquez, A. C., Schmidt, A. T., Chapman, S. B., Hotz, G. A., Yang, T. T., & Levin, H. S. (2012). Psychiatric disorders after pediatric traumatic brain injury: a prospective, longitudinal, controlled study. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *24*(4), 427-436.

McCarron, R. (2017). A survey of methods used for determining novel psychiatric research diagnoses in children and adolescents after Acquired Brain Injury, and their limitations. *Clinical Neuropsychiatry*, *14*(3), 195-208.

McCarron, R.H., Watson, S., & Gracey, F. (2019). What do kids with Acquired Brain Injury Want? Mapping neuropsychological rehabilitation goals to the International Classification of Functioning, Disability and Health. *Journal of the International Neuropsychological Society*, *25*(4), 403-412.

McKinlay, A., Grace, R., Horwood, J., Fergusson, D., & MacFarlane, M. (2009). Adolescent psychiatric symptoms following preschool childhood mild traumatic brain injury: evidence from a birth cohort. *The Journal of Head Trauma Rehabilitation*, *24*(3), 221-227.

Membride, H. (2016). Mental health: early intervention and prevention in children and young people. *British Journal of Nursing*, *25*(10), 552-557.

NHS England. (2013). *2013/2014 NHS Standard Contract for Paediatric Neurosciences: Neurorehabilitation.* NHS England.

Ornstein, T.J., Max, J.E., Schachar, R., Dennis, M., Barnes, M., Ewing-Cobbs, L., & Levin, H.S. (2013). Response inhibition in children with and without ADHD after traumatic brain injury. *Journal of Neuropsychology, 7*(1), 1-11.

Pastore, V., Galbiati, S., Recla, M., Colombo, K., Beretta, E., & Strazzer, S. (2018). Psychological and behavioural difficulties following severe TBI in adolescence: a comparison with a sample of peers with brain lesions of other origin and with a control group. *Brain Injury*, *32*(8), 1011-1020.

Ponsford, J. L., Downing, M. G., Olver, J., Ponsford, M., Acher, R., Carty, M., & Spitz, G. (2014). Longitudinal follow-up of patients with traumatic brain injury: outcome at two, five, and ten years post-injury. *Journal of Neurotrauma*, *31*(1), 64-77.

Prigatano, G. P., & Gupta, S. (2006). Friends after traumatic brain injury in children. *The Journal of Head Trauma Rehabilitation*, *21*(6), 505-513.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.

Schwartz, L., Taylor, H. G., Drotar, D., Yeates, K. O., Wade, S. L., & Stancin, T. (2003). Long-term behavior problems following pediatric traumatic brain injury: prevalence, predictors, and correlates. *Journal of Pediatric Psychology*, *28*(4), 251-263.

Simblett, S. K., Gracey, F., Ring, H., & Bateman, A. (2015). Measuring coping style following acquired brain injury: a modification of the Coping Inventory for Stressful Situations Using Rasch analysis. *British Journal of Clinical Psychology*, *54*(3), 249-265.

Sinopoli, K.J., Schachar, R., & Dennis, M. (2011). Traumatic brain injury and secondary attention-deficit/hyperactivity disorder in children and adolescents: the effect of reward on inhibitory control. *Journal of Clinical and Experimental Neuropsychology, 33*(7), 805-819.

Smith, J. E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205-231.

Smith, R.M., Linacre, J.M., and Smith, Jr., E.V. (2003). Guidelines for Manuscripts. *Journal of Applied Measurement*, *4*, 198-204.

Soo, C., Tate, R., & Brookes, N. (2014). Psychosocial adjustment following acquired brain injury in childhood and adolescence: Executive, behavioural and emotional contributions. *Brain Injury*, *28*(7), 906-914.

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical Child and Family Psychology Review*, *13*(3), 254-274.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, *57*(8), 1358-1362.

Wassenberg, R., Max, J. E., Lindgren, S. D., & Schatz, A. (2004). Sustained attention in children and adolescents after traumatic brain injury: Relation to severity of injury, adaptive functioning, ADHD and social background. *Brain Injury*, *18*(8), 751-764.

Wright, B. D. (1995). Diagnosing person misfit. *Rasch Measurement Transactions*, *9*(2), 430-431.

Yeates, K. O., Armstrong, K., Janusz, J., Taylor, H. G., Wade, S., Stancin, T., & Drotar, D. (2005). Long-term attention problems in children with traumatic brain injury. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(6), 574-584.

Yeates, K. O., Bigler, E. D., Dennis, M., Gerhardt, C. A., Rubin, K. H., Stancin, T., ... & Vannatta, K. (2007). Social outcomes in childhood brain disorder: a heuristic integration of social neuroscience and developmental psychology. *Psychological Bulletin*, *133*(3), 535.

Yeates, K. O., Gerhardt, C. A., Bigler, E. D., Abildskov, T., Dennis, M., Rubin, K. H., ... & Vannatta, K. (2013). Peer relationships of children with traumatic brain injury. *Journal of the International Neuropsychological Society*, *19*(5), 518-527.

Yeates, K. O., Taylor, H. G., Barry, C. T., Drotar, D., Wade, S. L., & Stancin, T. (2001). Neurobehavioral symptoms in childhood closed-head injuries: Changes in prevalence and correlates during the first year postinjury. *Journal of Pediatric Psychology*, *26*(2), 79-91.

Ylvisaker, M., Todis, B., Glang, A., Urbanczyk, B., Franklin, C., DePompei, R., Feeney, T., Maxwell, N. M., Pearson, S. M. A. & Tyler, J. S. (2001). Educating students with TBI: Themes and recommendations. *The Journal of Head Trauma Rehabilitation*, *16*(1), 76-93.

Table 1. Construct validity and internal reliability for the parent-reported Strengths and Difficulties subscales and the revised versions proposed in this study

Subscale	**Items**	**Rescoring pattern**	**Item-trait interaction**			**Reliability**		**Unidimensionality**	
			$\chi 2$	**$\chi 2$(df)**	***p* value**	**PSI**	**Cronbach's *α***	**Percentage of t-tests <.05**	**Lower bound 95% CI**
Emotional Symptoms	3, 8, 13, 16, 24	All items 012	14.95	10	0.13	0.65	0.78	0.00	-0.039
Emotional Symptoms Revised	8, 13, 16, 24	All items 012	9.16	8	0.33	0.66	0.81	0.93	-0.030
Conduct Problems	5, 7, 12, 18, 22	All items 012	19.33	10	0.036	0.67	0.78	0.81	-0.030
Conduct Problems Revised	5, 7, 12, 18, 22	Item 22 rescored 011, all other items 012	23.30	10	0.0097	0.66	0.77	0.82	-0.030
Hyperactivity	2, 10, 15, 21, 25	All items 012	15.75	10	0.11	0.63	0.76	2.42	0.11
Peer Problems	6, 11, 14, 19, 23	All items 012	13.38	10	0.20	0.60	*	1.68	-0.022
Prosocial	1, 4, 9, 17, 20	All items 012	13.23	10	0.34	0.69	*	0.83	0.031
Total Difficulties	2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 23, 24, 25	All items 012	68.42	40	0.0034	0.86	0.87	14.28	0.11
Total Difficulties Revised	2, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25	Items 12 and 22 rescored 011, all other items 012	54.40	36	0.025	0.86	0.87	14.89	0.11
Impact Supplement	Impact items 1, 2, 3, 4, 5	All items 0012	13.5	10	0.20	0.68	*	1.01	-0.033

PSI=Person Separation Index

CI=Confidence Interval

*=Cronbach's *α* not applicable due to missing data

Figure 1. Rasch models used in this study

The Masters Partial Credit Model

$$\ln\left(\frac{P_{nij}}{1-P_{nij-1}}\right)=\theta_n-\delta_{ij}$$

The Andrich Rating Scale Model

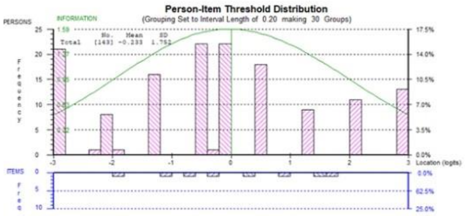
$$\ln\left(\frac{P_{nij}}{1-P_{nij-1}}\right)=\theta n-b_i-\tau_j$$

Where *P* is the probability of person *n* affirming item *i* with multiple response categories (*j)*. *Θ* is the overall level of mental health problem and *δ* is the specific degree of psychosocial strength or difficulty assessed by the item. τ is the threshold between adjacent response categories.

Figure 2. Person-item-threshold maps for the parent-reported Strengths and Difficulties Questionnaire

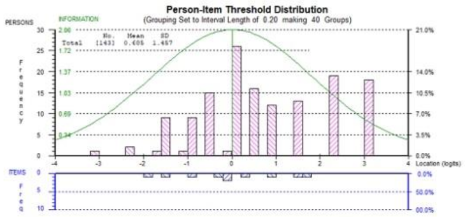
a) Revised Emotional Symptoms subscale

b) Revised Conduct Problems subscale



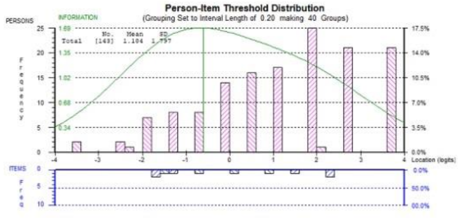
c) Hyperactivity subscale

d) Peer Problems subscale



e) Prosocial subscale

f) Revised Total Difficulties scale



g) Impact Supplement

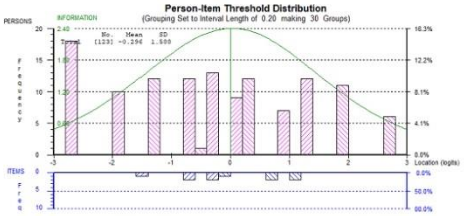
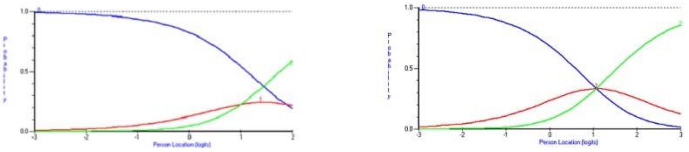


Figure 2. Person-item-threshold maps for the parent-reported Strengths and Difficulties Questionnaire

294x370mm (96 x 96 DPI)

Figure 3. Category probability curves for items with disordered thresholds in the parent-reported Strengths and Difficulties Questionnaire



a) Item 22 (Conduct Problems subscale analysis) "Steals from home, school or elsewhere"

b) Item 12 (Total Difficulties scale analysis) "Often fights with other children or bullies them"

c) Item 22 (Total Difficulties scale analysis) "Steals from home, school or elsewhere"

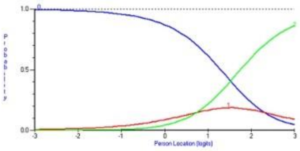


Figure 3. Category probability curves for items with disordered thresholds in the Strengths and Difficulties Questionnaire

387x255mm (96 x 96 DPI)

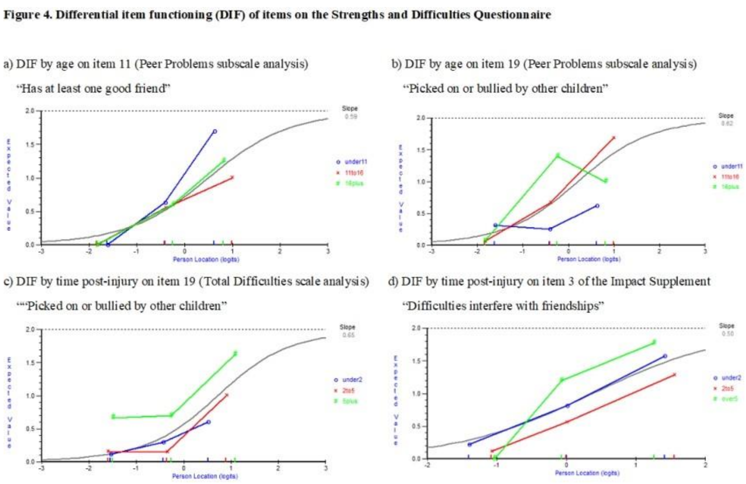


Figure 4. Differential item functioning (DIF) of items on the Strengths and Difficulties Questionnaire

a) DIF by age on item 11 (Peer Problems subscale analysis)

"Has at least one good friend"

b) DIF by age on item 19 (Peer Problems subscale analysis)

"Picked on or bullied by other children"

c) DIF by time post-injury on item 19 (Total Difficulties scale analysis)

""Picked on or bullied by other children"

d) DIF by time post-injury on item 3 of the Impact Supplement

"Difficulties interfere with friendships"

Figure 4. Differential item functioning (DIF) of items on the Strengths and Difficulties Questionnaire

387x255mm (96 x 96 DPI)