

The 2014 Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*) Part I:  
A Checklist for Dimensional Measurement of Selected DSM-5 Disorders

Laura Duncan, MA, PhD Student and Research Coordinator, McMaster University<sup>1,2</sup>

Katholiki Georgiades, PhD, Associate Professor, McMaster University<sup>1,2</sup>

Li Wang, MSc, PhD Student and Analyst, McMaster University<sup>1,2</sup>

Jinette Comeau, PhD, Assistant Professor, King's University College at Western University<sup>5</sup>

Mark A. Ferro, PhD, Assistant Professor, University of Waterloo<sup>6</sup>

Ryan J. Van Lieshout, MD, PhD, FRCPC, Assistant Professor, McMaster University<sup>1,2</sup>

Peter Szatmari, MD, MSc, FRCPS, Professor and Head, Centre for Addiction and Mental Health, Hospital for Sick Children, University of Toronto<sup>1,7</sup>

Kathryn Bennett, PhD, Professor, McMaster University<sup>1,4</sup>

Harriet L. MacMillan, MD, MSc, FRCPC, Professor, McMaster University<sup>1,2,3</sup>

Ellen L. Lipman, BSc, MD, FRCP(C), MSc, Professor and Director, McMaster University<sup>1,2</sup>

Magdalena Janus, PhD, Associate Professor, McMaster University<sup>1,2</sup>

Anna Kata, MA, MPH, Research Coordinator, McMaster University<sup>1,2</sup>

Michael H. Boyle, PhD, Professor Emeritus, McMaster University<sup>1,2</sup>

<sup>1</sup>Offord Centre for Child Studies; <sup>2</sup>Department of Psychiatry and Behavioural Neurosciences;

<sup>3</sup>Department of Paediatrics; <sup>4</sup>Department of Health Research Methods, Evidence and Impact (formerly Clinical Epidemiology and Biostatistics); McMaster University, 1280 Main St W, Hamilton, ON, L8S 4K1, Canada

<sup>5</sup>Department of Sociology, King's University College at Western University, 266 Epworth Ave, London, ON, N6A 2M3, Canada

<sup>6</sup>School of Public Health and Health Systems, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

<sup>7</sup>Department of Psychiatry, University of Toronto, 27 King's College Circle, Toronto, ON, M5S 1A1, Canada

**Correspondence** concerning this article should be addressed to: Laura Duncan, Offord Centre for Child Studies, McMaster University, 1280 Main St, West, MIP 201A, Hamilton Ontario, L8S 4K1. Email: [duncanlj@mcmaster.ca](mailto:duncanlj@mcmaster.ca) Phone: 1 905 525 9140 x21504 Fax: 905 574 6665

## Abstract

**Objectives.** To describe the development and psychometric properties of the 2014 Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*) for dimensional measurement of 7 disorders based on DSM-5 criteria.

**Methods.** Scale items were selected by agreement among 19 child psychologists and psychiatrists rating the correspondence between item descriptions and DSM-5 symptoms. Psychometric evaluation of the item properties and parent and youth scales came from a general population study of 10,802 children and youth aged 4 to 17 years in 6,537 families. Test-retest reliability data were collected from a subsample of 280 children and their parents who independently completed the *OCHS-EBS* checklist on two occasions 7 to 14 days apart. Structural equation modelling was used to assess internal and external convergent and discriminant validity—the latter tested against the Mini International Neuropsychiatric Interview for Children and Adolescents (*MINI-KID*).

**Results.** Confirmatory factor analyses exhibited adequate item-fit to all scales. Except for conduct disorder and youth-assessed separation anxiety disorder, internal (Cronbach's alpha) and test-retest reliability (Pearson's  $r$ ) for scale scores were 0.70 or above. Except for youth-assessed conduct disorder, the *OCHS-EBS* met criteria for internal and convergent and discriminant validity. When compared with the *MINI-KID*, the *OCHS-EBS* met criteria for external convergent and discriminant validity.

**Conclusions.** The *OCHS-EBS* provide reliable and valid dimensional measurement of seven DSM-5 disorders assessed by caregivers and youth in the general population. Part II describes use of the *OCHS-EBS* as a categorical (present/absent) measure of disorder.

**Keywords** Symptom checklist, Measurement, Structural equation modelling, Validity, Reliability, Child psychiatric disorder

## Introduction

Self-completed symptom checklists of child and adolescent psychiatric disorders are inexpensive to implement, pose little burden to respondents and can be administered in almost any setting to multiple informants (e.g., parents, teachers, and youth) using various modes of administration (e.g., in person, by mail, internet, telephone).<sup>1</sup> Many checklists have been developed to measure childhood psychopathology dimensionally including the Child Behavior Checklist (CBCL)<sup>2</sup> and the Strengths and Difficulties Questionnaire (SDQ).<sup>3</sup> However, these types of measures are limited in terms of (1) efficiency, (2) conceptualization and (3) versatility. First, the CBCL is long at over 100 items, while the SDQ is short but at the expense of coverage (it includes emotional symptoms, conduct problems and hyperactivity only). Second, no a priori attempt was made in the measurement development process to align items and syndromes with conceptualizations of disorder based on the Diagnostic and Statistical Manual 5<sup>th</sup> edition (DSM-5).<sup>4</sup> Third, developed prior to consensus on the practical and theoretical advantages of measuring psychiatric disorder as both dimensional and categorical phenomena,<sup>5-8</sup> there is no evidence that the scales associated with these measures, when converted to categorical measures of disorder (present/absent), are able to classify disorder as reliably and validly as structured interviews.<sup>9</sup> The Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*) address these limitations by balancing the number of items selected (burden) against comprehensive coverage of common disorders, selecting items according to DSM-5 disorder symptoms and serving both the needs of decision makers (categorical measures) and the pragmatics of measurement and analysis (dimensional measures). The current study focuses on the *OCHS-EBS* as dimensional measures of disorders and (1) describes how the development of these scales address the limitations of existing measures; and (2) presents the reliability and validity of these scales for measuring child psychiatric disorders as dimensional phenomena. A separate paper (Part II) evaluates the *OCHS-EBS* when used as a categorical (present/absent) measure of disorders.

## Development

The following practical requirements guided the development of the *OCHS-EBS*. *One*, in implementing the 2014 OCHS—a sequel to the original 1983 study<sup>10,11</sup>—we wanted scales to assess disorders commonly reported in general population surveys.<sup>12</sup> These included the following DSM-5 conditions: generalized anxiety disorder (GAD); separation anxiety disorder (SAD); major depressive disorder (MDD); social anxiety disorder (social phobia) (SP); attention-deficit hyperactivity disorder (ADHD); oppositional-defiant disorder (ODD); and conduct disorder (CD). *Two*, aware of declining response rates associated with the burden of participating in general population surveys,<sup>13</sup> we chose a completion time expected to fall within the tolerance of prospective respondents—7 to 10 minutes (about 50 items). *Three*, in measuring each disorder, we wished to achieve a similar standard of reliability and validity with the minimal number of items. This meant using clinical judgments as the basis for selecting items optimally matched with DSM-5<sup>4</sup> symptoms. Our primary focus was the development of a parent or caregiver-reported assessment for children aged 4 to 17 but we also evaluated an identical youth-reported assessment for youth aged 12 to 17.

The authors created a pool of 72 items by consensus to represent DSM-5 symptom criteria—64 taken from the Ontario Child Health Study-Revised (OCHS-R) scales<sup>14</sup> and 8 newly formulated to cover unrepresented symptoms. Nineteen child psychologists, psychiatrists and epidemiologists not involved in developing the item pool were asked to assess each item in relation to DSM-5 symptom criteria by independently (1) rating the extent to which its content

captured the meaning of its DSM-5 symptom analog; and (2) rank ordering the set of items associated with each scale in terms of how well they represented the core of each DSM-5 disorder. The item rating was scored as *1=no correspondence, 2=poor correspondence, could be interpreted to mean something else, 3=good correspondence, provides similar information and meaning and represents the symptom adequately, and 4=excellent match, provides almost the same information and meaning and represents the symptom very well*. The item ranking involved ordering the list of items associated with each DSM-5 disorder as to how well they represented the disorder overall.

To provide assurance that selected checklist items captured the operational meaning of each disorder (content validity), our criterion for selecting individual items was statistically significant agreement ( $P<0.015$  based on the sign test) achieved when 14 of 19 clinicians rated the item as providing (3) good or (4) excellent correspondence to its DSM-5 symptom analogue. When more than one item per symptom met this rating criterion, the one with a higher ranking was selected. Items not meeting the rating criteria were added by the development team if they were deemed highly representative of the disorder based on expert rankings. Fifty-five items met the criterion for consensus agreement among raters. Based on high rankings, the development team added 3 CD items ('Gets in many fights', 'Sets fires', 'Steals outside the home') and one SP item ('Doesn't like to be with people he/she doesn't know') from the item pool not meeting rating criteria for a total of 59 items.

## **Evaluation**

### **Methods**

#### **Participants**

This study uses data from the 2014 Ontario Child Health Study (OCHS),<sup>15</sup> an epidemiological study of children and youth aged 4 to 17 years old and their families, designed by researchers at McMaster University and conducted by Statistics Canada. Using the Canadian Child Tax Benefit file as the sampling frame, 15,796 dwellings were selected, 12,871 were eligible and 6,537 participated (50.8%). Dwellings were selected based on a complex 3-stage survey design that involved cluster sampling of residential areas and stratification by residency (urban, rural) and income (areas and households cross-classified by three levels of income (<20th; 20th to 80th; >80th percentiles). Within families, the primary parent/caregiver, their partner or spouse, and up to 4 children per family were interviewed resulting in 10,802 primary parent/caregiver reports on all youth aged 4 to 17 years old and 4,428 youth reports for youth aged 12 to 17 years old. To assess the reliability of study measures, a subsample of 180 caregivers and up to 2 of their children were re-interviewed 7 to 14 days after the initial interview. To obtain this subsample, Statistics Canada increased the number of dwellings chosen in selected urban clusters representing the three income strata and invited eligible families to participate until a total sample of 180 families was achieved. Interviewers provided a brief description of the study and booked consenting families. All families were interviewed at their homes by trained Statistics Canada interviewers. All study procedures, including consent and confidentiality requirements, were approved by the Chief Statistician at Statistics Canada and were conducted according to the *Statistics Act*.<sup>16</sup> Families were interviewed between October 2014 and October 2015. The sample analysis includes respondents with complete data on study measures—10,495 4 to 17 year olds (2.9% sample loss) and 3,945 youth aged 12 to 17 years old (10.9% sample loss).

## Concepts and Measures

### *ONTARIO CHILD HEALTH STUDY EMOTIONAL BEHAVIOURAL SCALES (OCHS-EBS) ITEMS*

Identical checklists of items from the item pool were completed by parents or caregivers of 4 to 17 year olds, and 12 to 17 year olds themselves as a self-administered paper (caregivers) or computerized (youth) questionnaire. Items were randomly ordered but in the same random order for both respondents. Respondents rated how well the statement describes the child or youth in the past 6 months as: '0=never or not true', '1=sometimes or somewhat true' and '2=often or very true'. Included in the analysis are respondents with no missing scale items which excluded only 0.75% of parent/caregivers and 0.9% of youth.

### *MINI INTERNATIONAL NEUROPSYCHIATRIC INTERVIEW FOR CHILDREN AND ADOLESCENTS (MINI-KID)*

Based on the adult MINI,<sup>17,18</sup> the *MINI-KID* is a standardized diagnostic interview that assesses DSM-IV-TR disorders in children and youth aged 6 to 17 years. Evaluated in 2 studies,<sup>19,20</sup> the *MINI-KID* exhibits good test-retest reliability ( $\kappa=0.56$  to  $0.87$ ) for mood, anxiety, substance use, ADHD, behavioural and eating disorders based on joint caregiver-child interviews and adequate agreement with another diagnostic interview.<sup>19</sup> The *MINI-KID* was administered separately to youth and caregivers.

In the 2014 OCHS, 7 disorder modules were administered independently to one randomly selected child per family and their caregiver ( $n=6,537$ ). The *MINI-KID* training given to Statistics Canada lay interviews included: (a) supervisor-led reading and review of an interviewer manual; (b) a guided training video on characteristics and symptom criteria of the included disorders and the *MINI-KID*, led by experienced researchers from McMaster University; (c) watching example video interviews during the training session; and (d) practice interviews among the interviewers. Interviewers were trained to ask the questions as worded; refrain from probing; encourage yes/no answers; and follow a protocol after "don't know" responses to ensure standardized administrations in accordance with procedures outlined by the *MINI-KID* authors. An interviewer dictionary provided standard definitions for terms and phrases used.

## Analysis

### Internal Validity

To evaluate empirically the 59 items remaining from scale development, we used confirmatory factor analysis (CFA) in Mplus 7.4<sup>21</sup> to confirm the expert item selection and assess the associations of the parent or caregiver-reported items with their hypothesized scales (internal factor structure). Confirmatory, as opposed to exploratory, factor analysis (EFA) was used as the number of factors being assessed and the organization of items within factors was already determined. However, correlations between items and scales were examined to determine if scale adjustments were required—a step that is typically part of an EFA procedure<sup>22</sup>. Based on Likert's method for summated rating scales,<sup>23</sup> we expected that items selected for each scale would represent each disorder as a unidimensional attribute, be associated with (load onto) their hypothesized scale at  $\geq 0.60$ <sup>24</sup> and provide adequate model fit to the observed data. Indicators of model fit and their criteria included: the comparative fit index (CFI: $>0.95$ ); and the root-mean-squared error of approximation (RMSEA: $<0.06$ ).<sup>25</sup> The chi-square test results of model fit are not used to assess model fit because large samples generate significant values even when there is satisfactory model fit.<sup>24</sup> Using the same CFA model fit criteria as above,<sup>26</sup> we expected measurement invariance (configural, metric and scalar) for each age group (age 4 to 11 and age 12 to 17) based on the caregiver report and for males and females

based on caregiver and youth reports. Configural measurement invariance indicates that the same items are associated with the same scales across all groups, metric invariance indicates factor loadings are similar across groups and scalar indicates that scale means are equivalent across groups.<sup>27</sup>

#### Internal Consistency and Test-retest Reliability

Internal consistency and test-retest reliability were expected to meet commonly accepted psychometric criteria<sup>28,29</sup> which include estimates  $\geq 0.70$  for both Cronbach's alpha (internal consistency) and Pearson's  $r$  (test-retest reliability).

#### Internal Convergent and Discriminant Validity

Building on the standard multitrait-multimethod (MTMM) approach to construct validation,<sup>30</sup> we used variance-based structural equation modelling to assess internal convergent and discriminant validity.<sup>21,31,32</sup> This method improves on the original MTMM approach by using objective criteria to evaluate construct validity and provides more sophisticated measurement of constructs.<sup>31</sup>

Convergent validity focuses on items that make up a scale and compares their shared variance with that scale (true measurement) in relation to their residual variance (measurement error); it is assessed using the Average Variance Extracted (AVE) and is demonstrated when the value of AVE is  $\geq 0.5$  indicating that at least 50% of the total variance in the items quantified by their factor loadings is explained by the scale.<sup>33,34</sup>

Discriminant validity focuses on association between items and their hypothesized scales in relation to their association with other scales in the set;<sup>35</sup> it is assessed by comparing the shared variance *within* each scale to the shared variance *between* scales and is demonstrated when the square root of AVE for a given scale is larger than the correlations between this scale and all others.<sup>34,36</sup> We expect some disorder overlap within individuals due to high rates of comorbidity,<sup>37,38</sup> (e.g., depression and anxiety<sup>39</sup>) and shared symptom profiles for some disorders (e.g., irritability and moodiness appear in ODD, MDD and GAD). As a result, the ability to discriminate between highly related or comorbid disorders will be reduced.<sup>40</sup>

#### External Convergent and Discriminant Validity

To evaluate the external convergent and discriminant validity of the scales empirically, we compared the *OCHS-EBS* with independent *MINI-KID* disorder assessments. First, we estimated point-biserial correlations between instruments for parent and youth assessments. We expect the correlations between instruments of the same disorder to be higher than the between-instrument correlations for non-similar disorders. Second, we implemented a similar MTMM CFA approach as used to evaluate our item selection. In our analysis here, we incorporated different informants (caregiver, youth), different instruments (*OCHS-EBS*, *MINI-KID*) and the disorders included in *OCHS-EBS*. The model consists of two factors—one representing internalizing disorder derived from GAD, SAD, MDD, SP; and the other, externalizing disorder derived from CD, ODD, ADHD—for each informant and instrument type (Figure 1 in the supplementary appendix). As done with the items, model fit was assessed using CFI and RMSEA. Evidence of convergent validity required the Average Variance Extracted (AVE) to be  $\geq 0.5$ ; and evidence of discriminant validity required the square root of AVE values to be larger than the inter-factor correlations among different disorder groupings assessed by the same or different informants or instruments. For example, discriminant validity is confirmed when the

square root of AVE for the caregiver checklist-assessed internalizing construct is larger than the inter-factor correlations between this construct and both the checklist and interview-assessed externalizing construct assessed by both the caregiver and youth.

## Results

Table 1 presents summary statistics for the 2014 OCHS study sample weighted by their probability of selection and the re-interviewed subsample (unweighted, as weights not available). The sample characteristics are almost identical although families had slightly higher incomes on average in the retest subsample. This was because families were sampled evenly across the three income strata resulting in low and high income families being overrepresented in the subsample.

### Internal Validity

Following confirmatory factor analysis with 59 eligible items, 7 selected items were dropped based on low factor loadings ( $<0.60$ ), high correlations with one or more different disorder scales or high correlations with other items (results available in Appendix). This left 52 items selected for 7 disorders: 48 meeting the criterion for symptom agreement among raters (14 or more of 19 raters) and 4 ranked highly as representing specific disorders (3 CD items: ‘Gets in many fights’, ‘Sets fires’, ‘Steals outside the home’ and one SP item: ‘Doesn’t like to be with people he/she doesn’t know’). All factor loadings exceeded 0.60; all models fit the data according to our criteria; and, except for CD, measurement invariance (configural, metric and scalar) of the factor structure was confirmed for all scales across sex (caregiver and youth report) and age groups (caregiver report) (results not shown).

### Internal Consistency and Test-retest Reliability

Table 2 displays the scale means and standard deviations by child sex, Cronbach’s alpha for internal consistency and test-retest reliabilities for caregiver report for ages 4 to 11 and 12 to 17, and youth report for ages 12 to 17. The scales are comprised of the same items across samples and informants. With the exception of youth-assessed CD and parent-assessed CD for ages 4 to 11, reliability estimates were all over 0.70 with one test-retest exception (youth-assessed SAD: 0.54). Average internal consistency was 0.80 for caregiver report for ages 4 to 11, 0.84 for caregiver report for ages 12 to 17 and 0.82 for youth report. Average test-retest reliability was 0.75, 0.79 and 0.74 for these three groups respectively.

### Internal Convergent and Discriminant Validity

Table 3 summarize the convergent and discriminant validity of the scales. Except for youth report CD, Average Variance Extracted (AVE) values for both caregiver and youth report scales were over 0.5 demonstrating convergent validity. Discriminant validity is established for a scale when the square root of AVE is larger than the correlations between this scale and all other scales in the measurement model. This was demonstrated in 35 out of 42 comparisons in the caregiver model and 25 out of 42 comparisons in the youth model. Discriminant validity test failures resulted from inter-factor correlations being larger than the square root of AVE for GAD (MDD), MDD (GAD, ODD), ODD (CD, MDD), CD (ODD) in the caregiver model; and for GAD (MDD), SAD (MDD), MDD (GAD, ADHD, ODD), SP (GAD, MDD), ADHD (MDD, ODD, CD), ODD (GAD, MDD, SAD, CD, ADHD) and CD (ODD, ADHD) in the youth model. Both models fit the data according to our criteria. Given the convergent validity failure of youth report CD, we repeated the analysis excluding CD. Convergent validity was established for the

remaining 6 scales and discriminant validity was demonstrated in 24 out of 30 cases (test failures resulted from higher square root AVE values than inter-factor correlations for GAD (MDD), MDD (GAD), ADHD (MDD, ODD) and ODD (MDD, ADHD)).

#### External Convergent and Discriminant Validity

Table 4 shows the correlations between the *OCHS-EBS* scale scores and *MINI-KID* disorder classifications for each informant. Correlations between instruments of the same disorder ranged from 0.37 to 0.59 for parents and 0.26 to 0.51 for youth. Between-instrument correlations for the same versus different disorders were higher in 81 of 84 comparisons. The exceptions were parent-assessed GAD (e.g., GAD-GAD=0.54; GAD-MDD=0.56), youth assessed MDD and youth assessed SP.

Table 5 summarizes the results of using CFA to model the 7 disorders scale scores by the *OCHS-EBS* and binary classifications by the *MINI-KID* for both caregiver and youth informants. AVE values in our model are over 0.5 providing evidence of convergent validity. Discriminant validity is established for all constructs in our model evidenced when the square root AVE value of a particular construct is larger than the inter-correlations between that construct and contrasting trait constructs based on both caregiver and youth report. CFI and RMSEA values provide evidence of good model fit, according to our criteria.

#### Discussion

This study presents the development and evaluation of the *OCHS-EBS*. From the initial pool of 72 items, 59 were selected for empirical evaluation—55 matched with specific DSM-5 symptoms by expert rating agreement and 4 with high disorder rankings to enhance coverage of CD and SP. Empirical evaluation based on CFA led to the removal of 7 items. The final scales consist of 52 items (6 GAD items, 7 SAD items, 9 MDD items, 5 SP items, 8 ADHD items, 6 ODD items, 11 CD items) that can be used to assess 7 DSM-5 disorders in children and youth aged 4 to 17 by summing responses to form a scale score for each disorder. For ease of use, selected items have been formatted into identical, alphabetically ordered parent and youth instruments together with scoring instructions and are provided as an electronic supplement.

Our scales performed well against the empirical standards of reliability and validity set in this study, with the exception of CD. Internal consistency reliability was less than  $\alpha=0.70$  for youth and caregiver (age 4 to 11) reported CD while test-retest reliability was less than  $r=0.70$  for youth reported CD. Although the internal convergent validity criterion was met for parent-reported CD, it was not met for youth-reported CD. Furthermore, in the youth model, CD was associated with many internal discriminant validity failures: excluding CD from the youth model reduced the number of internal discriminant validity test failures from 23 of 82 to 11 of 72.

The mixed success of CD was anticipated. Reliability is sample dependent, and scales measuring youth problem behaviour in general versus clinical populations will have lower means and variances, typically resulting in lower reliabilities as seen here.<sup>41</sup> This effect on reliability is compounded for CD because of the low prevalence of its symptoms. CD is an important part of the characterization of externalizing disorders along with ODD and ADHD. Despite its shortcomings as a scale, we recommend retaining the CD items in the *OCHS-EBS* to represent relatively rare and concerning behaviours.

Excluding youth-reported CD, CFA of individual items supported the internal convergent validity of the items selected to measure the disorders. However, there were a number of internal discriminant validity failures occurring between disorders that overlap with one another such as

GAD and MDD or ODD and ADHD. These failures reflect excessive overlap among individual child psychiatric disorders, particularly within the broad groupings of externalizing and internalizing disorders<sup>38,40</sup> which may be exacerbated in the *OCHS-EBS* by presenting the items in random order to reduce potential response bias.

Finally, evidence of external convergent and discriminant validity of the *OCHS-EBS* versus the *MINI-KID* for parent and youth informants was demonstrated for individual disorders in 81 out of 84 comparisons. Again, these exceptions occurred between disorders with similar characteristics (GAD & MDD and GAD & SP). When CFA was used to compare second order factors representing externalizing and internalizing disorders derived for each informant (parent, youth) and each instrument (*MINI-KID*, *OCHS-EBS*), evidence of external convergent and discriminant validity of the *OCHS-EBS* was demonstrated for individual disorders and their grouping into externalizing and internalizing constructs.

## Conclusion

In summary, this paper describes the development and properties of the *OCHS-EBS*, which are brief dimensional measures of 7 child psychiatric disorders based on DSM-5 criteria (GAD, SAD, MDD, SP, ADHD, ODD and CD). Following a rigorous item selection process based on expert clinician judgements, scales were evaluated using a large general population study from Ontario, Canada. Our evaluation indicates that the items and scales meet the psychometric requirements of validity and reliability for use as dimensional measures of child and adolescent psychiatric disorders, but that youth CD should be interpreted with caution. A variance-based structural equation model MTMM analysis provides evidence of both internal and external construct validity. This paper is based on a single general population study and further development and validation of the scales will be needed. Despite the large sample, this study does not include a clinical sample and it will be important to investigate the reliability and validity of this scale in other samples. The psychometric adequacy of these scales for measuring child and adolescent psychiatric disorders as a categorical phenomenon is the focus of a Part II companion paper.

**Data Access.** Data access available through Statistics Canada Research Data Centres.

**Conflict of Interest Disclosure.** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding.** This study was funded by research operating grant 125941 from the Canadian Institutes of Health Research (CIHR), Health Services Research Grant 8-42298 from the Ontario Ministry of Health and Long-Term Care (MOHLTC) and from funding from MOHLTC, the Ontario Ministry of Children and Youth Services and the Ontario Ministry of Education. Dr. Boyle is supported by CIHR Canada Research Chair in the Social Determinants of Child Health; Dr. Georgiades by a CIHR New Investigator Award and the David R. (Dan) Offord Chair in Child Studies; Dr. Ferro by the CIHR Canada Research Chair in Youth Mental Health and an Early Researchers Award from the Ontario Ministry of Research, Innovation and Science; and Dr. MacMillan by the Chedoke Health Chair in Child Psychiatry.

## References

1. Myers K, Winters NC. Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *J Am Acad Child Adolesc Psychiatry*. 2002;41(2):114–122.
2. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington (VT): Department of Psychiatry, University of Vermont; 1991.
3. Goodman R. Psychometric properties of the Strengths and Difficulties Questionnaire. *J Am Acad Child Adolesc Psychiatry*. 2001;40(11):1337–45.
4. American Psychiatric Association. Diagnostic and statistical manual of mental Disorders. 5th ed. Arlington (VA): American Psychiatric Publishing; 2013.
5. Coghill D, Sonuga-Barke EJS. Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders – implications of recent empirical study. *J Child Psychol Psychiatry*. 2012;53(5):469–489.
6. Kraemer HC, Noda A, O’Hara R. Categorical versus dimensional approaches to diagnosis: methodological challenges. *Journal of Psychiatric Research*. 2004;38(1):17–25.
7. Markon KE, Chmielewski M, Miller CJ. The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychol Bull*. 2011;137(5):856–879.
8. Pickles A, Angold A. Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Dev Psychopathol*. 2003;15(3):529–551.
9. Boyle MH, Duncan L, Georgiades K et al. Classifying child and adolescent psychiatric disorder by problem checklists and standardized interviews. *Int J Methods Psychiatr Res*. 2016;26(4):e1544 .
10. Boyle, M, Offord DR, Hofmann HF, et al. Ontario Child Health Study: I. Methodology. *Arch Gen Psychiatry*. 1987;44(9):826–831.
11. Offord DR, Boyle MH, Szatmari P, et al. Ontario Child Health Study: II. Six-month prevalence of disorder and rates of service utilization. *Arch Gen Psychiatry*. 1987;44(9):832–836.
12. Polanczyk GV, Salum GA, Sugaya LS, et al. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J Child Psychol Psychiatry*. 2015;56(3):345–365.
13. Bladon TL. The downward trend of survey response rates: Implications and considerations for evaluators. *Can J Program Eval*. 2010;24(2):131–156.
14. Boyle MH, Offord DR, Racine YA, et al. Evaluation of the revised Ontario Child Health Study Scales. *J Child Psychol Psychiatry*. 1993;34:189–213.
15. Statistics Canada. 2014 Ontario Child Health Study (master file). Statistics Canada. Survey No. 3824. 2017.
16. Government of Canada. Statistics Act. 2018. Retrieved from <http://laws-lois.justice.gc.ca/eng/acts/S-19/index.html> on June 7th, 2017.
17. Lecrubier Y, Sheehan DV, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry*. 1997;12(5):224–231.
18. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59:22–33.
19. Sheehan DV, Sheehan KH, Shytle RG, et al. Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *J Clin*

Psychiatry. 2010;71:313–326.

20. Duncan L, Georgiades K, Wang L, et al. Psychometric evaluation of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Psychol Assess*. 2017;30(7):916–928.

21. Muthén LK, Muthén BO. *Mplus User's Guide*. 6th ed. Los Angeles (CA): Muthén & Muthén; 2016.

22. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeduc Assess*. 2011;29(4):304–21.

23. Likert RA. A technique for the measurement of attitudes *Arch Psychol (Chic)*. 1932;140:1–55.

24. MacCallum RC, Widaman KF, Zhang S, et al. Sample size in factor analysis. *Psychol Methods*. 1999;4(1):84–99.

25. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1–55.

26. Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct Equ Modeling*. 2007;14(3):464–504.

27. Millsap RE. *Statistical approaches to measurement invariance*. New York (NY): Routledge; 2012.

28. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ*, 1997;314(7080):572.

29. DeVellis RF. *Scale development: Theory and applications*. Thousand Oaks (CA): Sage; 2012.

30. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959;56(2):81–105.

31. Andrasik F, editor. *Comprehensive handbook of personality and psychopathology, adult psychopathology (Vol. 2)*. New Jersey (NJ): John Wiley & Sons; 2005.

32. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci*. 2015;43(1):115–135.

33. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res*, 1981;39–50.

34. Hair JF, Anderson RE, Babin BJ, et al. *Multivariate data analysis: A global perspective (Vol. 7)*. New Jersey (NJ): Pearson; 2010.

35. Brown TA. *Confirmatory factor analysis for applied research*. New York (NY): Guilford Press; 2006.

36. Carter SR. Using confirmatory factor analysis to manage discriminant validity issues in social pharmacy research. *Int J Clin Pharm*. 2016;38(3):731–737.

37. Krueger RF, Markon KE. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annu Rev Clin Psychol*. 2006;27(2):111–33.

38. Clark LA, Watson D, Reynolds S. Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annu Rev Psychol*. 1995;46(1):121–53.

39. Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. *Annu Rev Psychol*. 1998;49(1):377–412.

40. Sher KJ, Trull TJ. Methodological issues in psychopathology research. *Annu Rev Psychol*. 1996;47(1):371–400.

41. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. New York(NY): Oxford University Press; 2014.

Table 1. Sample characteristics

	<i>2014 OCHS study sample<sup>a</sup></i>	<i>Retest sample</i>
<b>Youth</b>		
<i>n</i>	10,802	280
Age, <i>M</i> (SD)	10.63 (4.07)	10.11 (4.16)
Male, %	51.3	49.3
<b>Parent/Caregiver</b>		
<i>n</i>	6,537	180
Age, <i>M</i> (SD)	41.54 (7.20)	41.39 (6.87)
Male, %	11.8	16.7
<b>Family</b>		
<i>n</i>	6,537	180
Household Income <i>M</i> (SD)	\$100,500 (\$162,600)	\$114,000 (\$94,400)
Single parent, %	20.6	17.2

<sup>a</sup>Weighted according to the probability of selection

Table 2. Weighted variability and reliability of the 2014 OCHS scales by gender and age

	All	Mean (SD) Male	Female	Internal Consistency ( $\alpha$ )	Test-retest reliability ( $r$ ) <sup>a</sup>
Parent/Caregiver Report (ages 4 to 11) $n=6,203$					$n=148$
GAD	1.40 (2.03)	1.41 (2.04)	1.38 (2.02)	.81	.73
SAD	1.29 (2.04)	1.21 (1.99)	1.36 (2.08)	.80	.77
MDD	1.11 (1.70)	1.24 (1.84)	0.97 (1.52)	.70	.78
SP	1.61 (1.93)	1.53 (1.91)	1.70 (1.95)	.81	.70
ADHD	2.83 (3.21)	3.38 (3.43)	2.24 (2.83)	.87	.76
ODD	1.56 (1.98)	1.82 (2.16)	1.28 (1.73)	.79	.77
CD	0.33 (0.89)	0.43 (1.01)	0.22 (0.73)	.62	.71
Parent/Caregiver Report (ages 12 to 17) $n=4,292$					$n=105$
GAD	1.89 (2.38)	1.75 (2.34)	2.04 (2.42)	.85	.84
SAD	0.72 (1.59)	0.65 (1.51)	0.81 (1.67)	.81	.79
MDD	1.87 (2.55)	1.72 (2.48)	2.03 (2.61)	.83	.75
SP	1.78 (2.25)	1.70 (2.30)	1.87 (2.20)	.86	.78
ADHD	2.28 (2.93)	2.71 (3.19)	1.81 (2.54)	.87	.87
ODD	1.68 (2.22)	1.80 (2.31)	1.54 (2.10)	.84	.70
CD	0.44 (1.36)	0.52 (1.55)	0.36 (1.11)	.80	.82
Youth Report (ages 12 to 17) $n=3,925$					$n=96$
GAD	3.26 (3.05)	2.67 (2.75)	3.90 (3.23)	.86	.78
SAD	1.95 (2.49)	1.64 (2.21)	2.29 (2.72)	.79	.54
MDD	2.86 (3.32)	2.30 (2.75)	3.46 (3.76)	.85	.78
SP	3.08 (2.69)	2.75 (2.57)	3.44 (2.78)	.84	.78
ADHD	3.79 (3.19)	3.85 (3.26)	3.71 (3.11)	.81	.74
ODD	2.07 (2.20)	2.05 (2.16)	2.08 (2.24)	.76	.82
CD	0.90 (1.50)	0.97 (1.58)	0.83 (1.42)	.66	.60

<sup>a</sup>All estimates  $p < 0.01$

Note. 2014 OCHS Sample weighted based on the probability of selection. Retest subsample unweighted.

ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, GAD=Generalized Anxiety Disorder, MDD=Major Depressive Disorder, ODD=Oppositional Defiant Disorder, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia)

Table 3. Weighted scale AVE values, inter-factor correlations and fit indices for confirmatory factor analysis by informant

	Average variance extracted (AVE) ( $\sqrt{\text{AVE}}$ )	Inter-Factor Correlations					
		GAD	SAD	MDD	SP	ADHD	ODD
Parent/caregiver informant	GAD	0.67 (0.82)					
	SAD	0.67 (0.82)	0.68				
	MDD	0.59 (0.77)	0.83	0.57			
	SP	0.72 (0.85)	0.65	0.56	0.64		
	ADHD	0.64 (0.80)	0.63	0.48	0.68	0.43	
	ODD	0.64 (0.80)	0.67	0.49	0.80	0.51	0.76
	CD	0.61 (0.78)	0.50	0.30	0.67	0.33	0.72
Model Fit Indices							
$\chi^2$ (df) <sup>a</sup>	2264.367 (df=1253)	$P < 0.001$					
CFI <sup>b</sup>	0.943						
RMSEA <sup>c</sup>	0.009						
	Average variance extracted (AVE) ( $\sqrt{\text{AVE}}$ )	Inter-Factor Correlations					
		GAD	SAD	MDD	SP	ADHD	ODD
Youth informant	GAD	0.70 (0.84)					
	SAD	0.52 (0.72)	0.71				
	MDD	0.57 (0.76)	0.92	0.75			
	SP	0.60 (0.77)	0.77	0.66	0.74		
	ADHD	0.50 (0.70)	0.68	0.61	0.78	0.69	
	ODD	0.51 (0.71)	0.72	0.72	0.78	0.70	0.88
	CD	0.48 (0.69)	0.50	0.56	0.63	0.43	0.71
Model Fit Indices							
$\chi^2$ (df) <sup>a</sup>	3808.501 (df=1253)	$P < 0.001$					
CFI <sup>b</sup>	0.970						
RMSEA <sup>c</sup>	0.014						

<sup>a</sup> Chi-Square test

<sup>b</sup> Comparative Fit Index

<sup>c</sup> Root Mean Squared Error of Approximation

Note: ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, GAD=Generalized Anxiety Disorder, MDD=Major

Depressive Disorder, ODD=Oppositional Defiant Disorder, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia)

Table 4. Multi-trait, multi-method matrix showing point-biserial correlations between the *OCHS-EBS* scale scores and *MINI-KID* disorder classifications by informant.<sup>a</sup>

Note: ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, GAD=Generalized Anxiety Disorder, MDD=Major Depressive Disorder, MINI-KID=Mini International Neuropsychiatric Interview for Children and Adolescents interview, OCHS-

Method	Trait	<i>MINI-KID-P</i>							<i>MINI-KID-Y</i>								
		GAD	SAD	MDD	SP	ADHD	ODD	CD	GAD	SAD	MDD	SP	ADHD	ODD	CD		
<i>OCHS-EBS-P</i>	GAD	.54															
	SAD	.37	.37														
	MDD	.56	.25	.59													
	SP	.37	.14	.29	.48												
	ADHD	.33	.12	.33	.34	.48											
	ODD	.31	.12	.36	.29	.34	.51										
	CD	.24	.09	.27	.24	.30	.47	.47									
<i>OCHS-EBS-Y</i>	GAD									.51							
	SAD									.28	.26						
	MDD									.45	.33	.41					
	SP									.33	.17	.23	.32				
	ADHD									.28	.19	.23	.19	.29			
	ODD									.28	.21	.25	.23	.28	.36		
	CD									.19	.24	.21	.18	.24	.35	.37	

EBS=Ontario Child Health Study Emotional Behavioural Scales checklist, ODD=Oppositional Defiant Disorder, P=parent, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia), Y=youth

<sup>a</sup> All correlations  $P < 0.01$

Table 5. Weighted standardized factor loadings, AVE values, inter-factor correlations and fit indices for confirmatory factor analysis – parent/caregiver and youth informant

Model	Parent/Caregiver		Youth				
	Standardized factor loadings (error variance)	Average variance extracted (AVE) ( $\sqrt{\text{AVE}}$ )	Standardized factor loadings (error variance)	Average variance extracted (AVE) ( $\sqrt{\text{AVE}}$ )			
<i>OCHS-Int</i>		0.53 (0.73)		0.58 (0.76)			
GAD	0.79 (0.38)		0.81 (0.34)				
SAD	0.57 (0.68)		0.59 (0.65)				
MDD	0.88 (0.23)		0.93 (0.14)				
SP	0.62 (0.62)		0.67 (0.55)				
<i>OCHS-Ext</i>		0.61 (0.78)		0.58 (0.76)			
ADHD	0.60 (0.64)		0.62 (0.62)				
ODD	0.89 (0.21)		0.86 (0.26)				
CD	0.82 (0.33)		0.79 (0.38)				
<i>MINI-Int</i>		0.71 (0.85)		0.78 (0.88)			
GAD	0.94 (0.12)		0.90 (0.19)				
SAD	0.72 (0.48)		0.86 (0.26)				
MDD	0.88 (0.23)		0.91 (0.17)				
SP	0.82 (0.33)		0.87 (0.24)				
<i>MINI-Ext</i>		0.84 (0.91)		0.85 (0.92)			
ADHD	0.98 (0.04)		0.99 (0.02)				
ODD	0.98 (0.04)		0.92 (0.15)				
CD	0.89 (0.21)		0.86 (0.26)				
<b>Inter-Factor Correlations<sup>a</sup></b>							
	P- <i>OCHS-Int</i>	P- <i>OCHS-Ext</i>	P- <i>MINI-Int</i>	P- <i>MINI-Ext</i>	Y- <i>OCHS-Int</i>	Y- <i>OCHS-Ext</i>	Y- <i>MINI-Int</i>
P- <i>OCHS-Ext</i>	0.73						
P- <i>MINI-Int</i>	0.76	0.48					
P- <i>MINI-Ext</i>	0.47	0.71	0.57				
Y- <i>OCHS-Int</i>	0.43	0.27	0.48	0.23			
Y- <i>OCHS-Ext</i>	0.43	0.54	0.45	0.49	0.75		
Y- <i>MINI-Int</i>	0.40	0.33	0.72	0.34	0.70	0.49	

Y-MINI-Ext	0.35	0.51	0.52	0.77	0.33	0.62	0.61
<b>Model Fit Indices</b>							
$\chi^2$ (df) <sup>b</sup>	745.799 (df=323)		P<0.001				
CFI <sup>c</sup>	0.953						
RMSEA <sup>d</sup>	0.011						

<sup>a</sup> These correlations correspond to the paths identified in Figure 1 in the supplementary appendix. M=MINI-KID, O=OCHS-EBS,

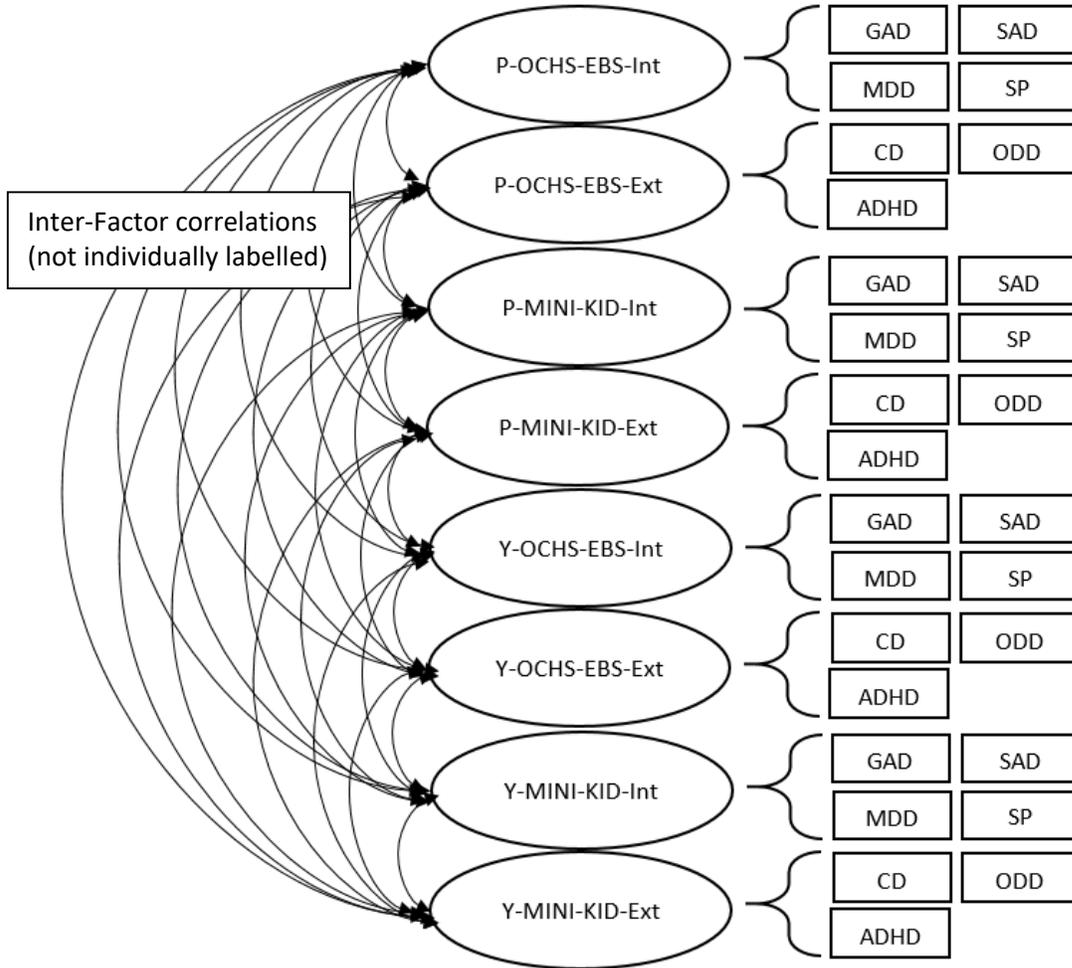
<sup>b</sup> Chi-Square test

<sup>c</sup> Comparative Fit Index

<sup>d</sup> Root Mean Squared Error of Approximation

Note: ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, Ext=Externalizing, GAD=Generalized Anxiety Disorder, Int= Internalizing, MDD=Major Depressive Disorder, MINI=Mini International Neuropsychiatric Interview for Children and Adolescents interview, OCHS=Ontario Child Health Study Emotional Behavioural Scales checklist, ODD=Oppositional Defiant Disorder, P=parent, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia), Y=youth

Figure 1. Multitrait-multimethod confirmatory factor analysis model for 8 factor model of internalizing and externalizing latent factors assessed using the OCHS-EBS checklist and MINI-KID by parent/caregiver and youth informants. The arrows on the left correspond to inter-factor correlations; values for which appear at the bottom of Table 5.



Note. ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, Ext=Externalizing, GAD=Generalized Anxiety Disorder, Int= Internalizing, MDD=Major Depressive Disorder, MINI-KID=Mini International Neuropsychiatric Interview for Children and Adolescents interview, OCHS-EBS=Ontario Child Health Study Emotional Behavioural Scales checklist, ODD=Oppositional Defiant Disorder, P=parent, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia), Y=youth

## Appendix

Table 1. Factor loadings for 52 OCHS-EBS items

Disorder	Item	Factor loading
GAD	Too fearful or anxious	.83
	Worries about doing better at things	.70
	Finds it hard to stop worrying	.74
	Anxious or on edge	.87
	Nervous, high-strung or tense	.83
	When anxious, his/her mind goes blank	.85
SAD	Overly upset when leaving loved ones	.92
	Worries that bad things will happen to loved ones	.86
	Worries that something bad will cause separation from loved ones	.88
	Avoids school because of fear of separation from loved ones	.95
	Scared to go to sleep without parents being near	.76
	Has nightmares about being separated from loved ones	.87
	Complains of feeling sick before separating from loved ones	.90
MDD	Unhappy, sad or depressed	.81
	Gets no pleasure from usual activities	.86
	Has trouble enjoying self	.88
	Changes in appetite	.71
	Trouble sleeping	.80
	Overtired or lacks energy	.67
	Feels worthless or inferior	.89
	Deliberately harms self or attempts suicide	.96
Talks about killing self (youth item: thinks about killing self)	.96	
SP	Doesn't like to be with people he/she doesn't know <sup>a</sup>	.83
	Afraid of doing things in front of others	.85
	Avoids social situations	.95
	Is nervous with people he/she doesn't know	.86
	Gets anxious about meeting new people	.88
ADHD	Makes careless mistakes	.69
	Can't concentrate, can't pay attention for long	.82
	Fails to finish things he/she starts	.88
	Distractible, has trouble sticking to any activity	.90
	Fidgets	.83
	Can't stay seated when required to do so	.85
	Impulsive or acts without thinking	.84
Has difficulty awaiting turn in games or groups	.84	
ODD	Loses temper	.86
	Argues a lot with adults	.86
	Blames others for own mistakes	.81
	Easily annoyed by others	.85
	Angry and resentful	.92
CD	Gets back at people	.91
	Cruelty, bullying or meanness to others	.82
	Gets in many fights <sup>a</sup>	.78

Uses weapons when fighting	.81
Has been physically cruel to others	.87
Destroys things belonging to his/her family or other children	.72
Has broken into someone else's house, building or car	.87
Sets fires <sup>a</sup>	.85
Steals outside the home <sup>a</sup>	.81
Stays out at night despite being told not to	.71
Runs away from home	.85
Truancy; skips school	.70

---

<sup>a</sup> Item did not meet rating criteria but added based on high clinician ranking indicating a core disorder symptom. Based on CFA with 59 items, seven items were dropped due to factor loading <0.60 (CD items: 'Cruelty to animals', 'Has mugged people'), high correlation with a different scale (GAD item: 'Moody or irritable', MDD item: 'Has difficulty making decisions', ADHD item: 'Can't sit still, restless or hyperactive'), high correlation with another item (GAD item: 'When anxious, he/she has disturbed sleep', ODD item: 'Defiant, talks back to adults')

Note. ADHD=Attention-Deficit Hyperactivity Disorder, CD=Conduct Disorder, GAD=Generalized Anxiety Disorder, MDD=Major Depressive Disorder, OCHS-EBS=Ontario Child Health Study Emotional Behavioural Scales checklist, ODD=Oppositional Defiant Disorder, SAD=Separation Anxiety Disorder, SP=Social Anxiety Disorder (Social Phobia)