

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

The Canadian Journal of Psychiatry /
La Revue Canadienne de Psychiatrie

1-11

© The Author(s) 2018

Article reuse guidelines:


sagepub.com/journals-permissions

DOI: 10.1177/0706743718808250

TheCJP.ca | LaRCP.ca



Échelles émotionnelles comportementales de l'Étude sur la santé des jeunes ontariens (EEC-ESJO) de 2014, première partie : une liste de vérification pour la mesure dimensionnelle de troubles choisis du DSM-5

Laura Duncan, MA^{1,2,3} , Katholiki Georgiades, PhD^{1,2}, Li Wang, MA^{1,2,3},
Jinette Comeau, PhD⁴, Mark A. Ferro, PhD⁵,
Ryan J. Van Lieshout, MD, PhD, FRCPC^{1,2}, Peter Szatmari, MD, FRCPS^{1,6},
Kathryn Bennett, PhD^{1,3}, Harriet L. MacMillan, MD, FRCPC^{1,2,7},
Ellen L. Lipman, MD, FRCPC^{1,2}, Magdalena Janus, PhD^{1,2},
Anna Kata, MA, MPH^{1,2}, and Michael H. Boyle, PhD^{1,2}

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 criteria from the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*.

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/caregiver and youth scales came from a general population study of 10,802 children and youth aged 4 to 17 years in 6537 families. Test-retest reliability data were collected from a subsample of 280 children and their caregivers who independently completed the OCHS-EBS checklist on 2 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 α) 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. Compared with the MINI-KID, the OCHS-EBS met criteria for external convergent and discriminant validity.

¹ Offord Centre for Child Studies, McMaster University, Hamilton, Ontario

² Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario

³ Department of Health Research Methods, Evidence and Impact (formerly Clinical Epidemiology and Biostatistics), McMaster University, Hamilton, Ontario

⁴ Department of Sociology, King's University College at Western University, London, Ontario

⁵ School of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario

⁶ Department of Psychiatry, Centre for Addiction and Mental Health, Hospital for Sick Children, University of Toronto, Toronto, Ontario

⁷ Department of Paediatrics, McMaster University, Hamilton, Ontario

Corresponding Author:

Laura Duncan, MA, Offord Centre for Child Studies, McMaster University, 1280 Main St. West, MIP 201A, Hamilton, Ontario L8S 4K1, Canada.

Email: duncanlj@mcmaster.ca

Conclusions: The OCHS-EBS provide reliable and valid dimensional measurement of 7 *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.

Abrégé

Objectifs : Décrire le développement des propriétés psychométriques des échelles émotionnelles comportementales de l'Étude sur la santé des jeunes ontariens (EEC-ESJO) de 2014 pour la mesure dimensionnelle de 7 troubles basés sur les critères du *DSM-5*.

Méthodes : Les items des échelles ont été choisis d'un commun accord par 19 psychologues et psychiatres pour enfants qui ont évalué la correspondance entre les descriptions des items et les symptômes du *DSM-5*. L'évaluation psychométrique des propriétés des items et des échelles des parents et des adolescents provenait d'une étude dans la population générale de 10 802 enfants et adolescents âgés de 4 à 17 ans et issus de 6 537 familles. Les données de fiabilité test-retest ont été recueillies d'un sous-ensemble de 280 enfants et leurs parents qui ont indépendamment répondu à la liste de vérification des EEC-ESJO à deux occasions, entre 7 à 14 jours d'intervalle. La modélisation par équation structurelle a servi à évaluer la validité convergente et discriminante interne et externe—la validité discriminante a été ensuite testée par rapport à la mini-entrevue neuropsychiatrique internationale pour enfants et adolescents (MINI-KID).

Résultats : Des analyses factorielles confirmatoires ont montré un ajustement des items adéquat dans toutes les échelles. À l'exception du trouble des conduites et du trouble d'anxiété de séparation évalué par les adolescents, la fiabilité interne (α de Cronbach) et de test-retest (r de Pearson) pour les scores aux échelles était de 0,70 ou plus. À l'exception du trouble des conduites évalué par les adolescents, les EEC-ESJO satisfaisaient aux critères de la validité convergente et discriminante interne. Comparées avec la MINI-KID, les EEC-ESJO satisfaisaient aux critères de la validité convergente et discriminante externe.

Conclusions : Les EEC-ESJO offrent une mesure dimensionnelle fiable et valide de sept troubles du *DSM-5* évalués par des parents et des adolescents de la population générale. La deuxième partie décrit l'utilisation des EEC-ESJO à titre de mesure catégorique (présent/absent) d'un trouble.

Keywords

symptom checklist, measurement, structural equation modelling, validity, reliability, child psychiatric disorder

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).¹ Many checklists have been developed to measure childhood psychopathology dimensionally, including the Child Behavior Checklist (CBCL)² and the Strengths and Difficulties Questionnaire (SDQ).³ 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 fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*.⁴ Third, developed prior to consensus on the practical and theoretical advantages of measuring psychiatric disorder as both dimensional and categorical phenomena,⁵⁻⁸ 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.⁹ 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 addresses 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 article (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^{10,11}—we wanted scales to assess disorders commonly reported in general population surveys.¹² 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,¹³ 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 judgements as the basis for selecting items optimally matched with *DSM-5*⁴ 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) scales¹⁴ 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 analogue 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 1 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 1 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),¹⁵ an epidemiological study of children and youth aged 4 to 17 years 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 6537 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 3 levels of income: <20th, 20th to 80th, and >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 children aged 4 to 17 years and 4428 youth reports for youth aged 12 to 17 years. To assess the reliability of study measures, a subsample of 180 caregivers and up to 2 of their children were reinterviewed 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 3 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.¹⁶ Families were interviewed between October 2014 and October 2015. The sample for analysis includes respondents with complete data on study measures—10,495 4- to 17-year-olds (2.9% sample loss) and 3945 youth aged 12 to 17 years (10.9% sample loss).

Concepts and measures

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,^{17,18} 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,^{19,20} the MINI-KID

exhibits good test-retest reliability ($\kappa = 0.56$ to 0.87) for mood, anxiety, substance use, ADHD, and behavioural and eating disorders based on joint caregiver-child interviews and adequate agreement with another diagnostic interview.¹⁹ The MINI-KID was administered separately to youth and caregivers.

In the 2014 OCHS, 7 disorder modules were administered independently to 1 randomly selected child per family and his or her caregiver ($n = 6537$). 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 7.4²¹ to confirm the expert item selection and assess the associations of the caregiver-reported items with their hypothesized scales (internal factor structure). CFA, 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.²² Based on Likert's method for summated rating scales,²³ 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 ≥ 0.60 ,²⁴ 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).²⁵ The χ^2 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.²⁴ Using the same CFA model fit criteria as above,²⁶ we expected measurement invariance (configural, metric, and scalar) for each age group (ages 4 to 11 and ages 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.²⁷

Internal consistency and test-retest reliability. Internal consistency and test-retest reliability were expected to meet commonly accepted psychometric criteria,^{28,29} which include estimates ≥ 0.70 for both Cronbach's α (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,³⁰ we used variance-based structural equation modelling to assess internal convergent and discriminant validity.^{21,31,32} This method improves on the original MTMM approach by using objective criteria to evaluate construct validity and provides more sophisticated measurement of constructs.³¹

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 ≥ 0.5 , indicating that at least 50% of the total variance in the items quantified by their factor loadings is explained by the scale.^{33,34}

Discriminant validity focuses on associations between items and their hypothesized scales in relation to their association with other scales in the set³⁵; 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.^{34,36} We expect some disorder overlap within individuals due to high rates of comorbidity^{37,38} (e.g., depression and anxiety³⁹) 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.⁴⁰

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 caregiver and youth assessments. We expect the correlations between instruments of the same disorder to be higher than the between-instrument correlations for nonsimilar 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 2 factors—one representing internalizing disorder derived from GAD, SAD, MDD, and SP and the other representing externalizing disorder derived from CD, ODD, and ADHD—for each informant and instrument type (Figure 1). As done with the items, model fit was assessed using CFI and RMSEA. Evidence of convergent validity required the AVE to be ≥ 0.5 , and evidence of discriminant validity

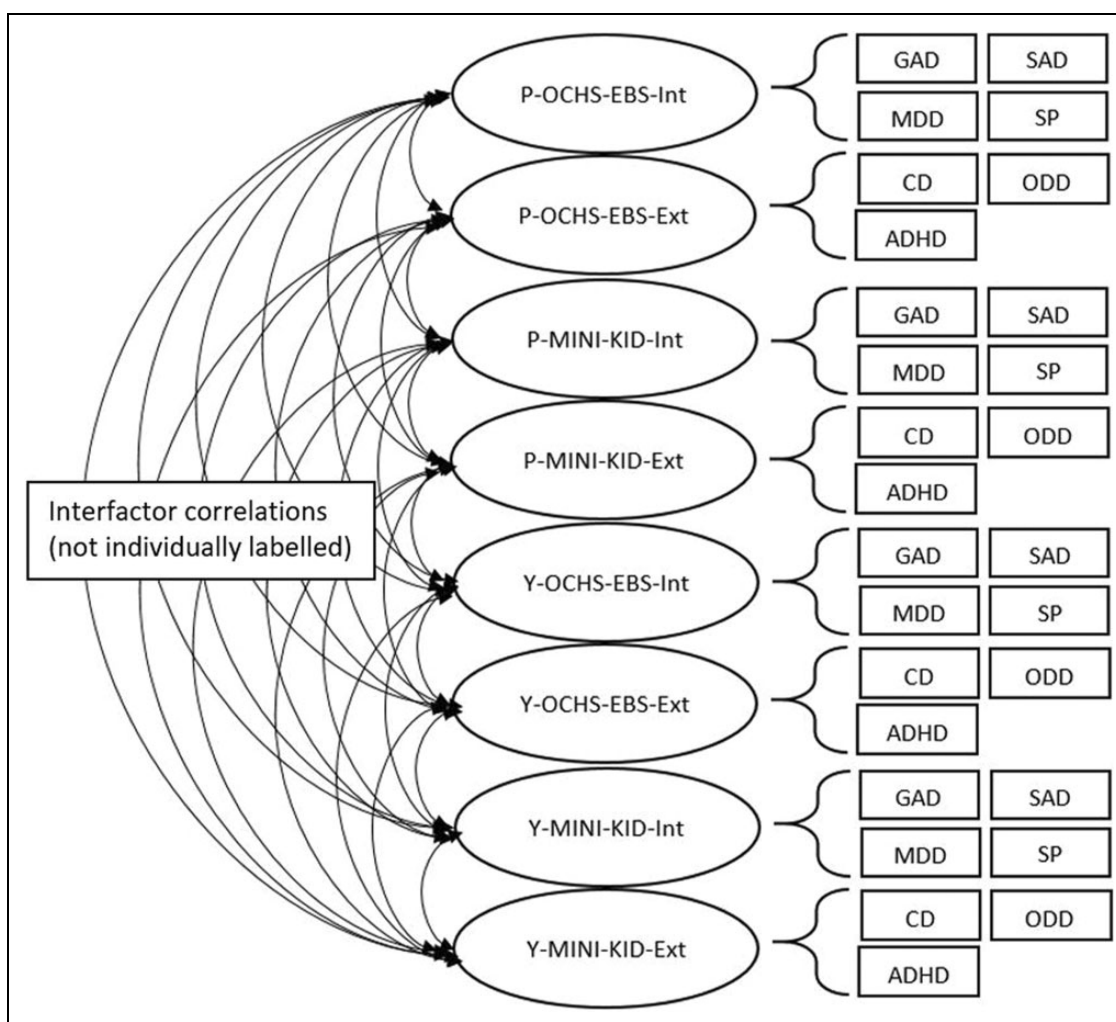


Figure 1. Multitrait-multimethod confirmatory factor analysis model for 8-factor model of internalizing and externalizing latent factors assessed using the Ontario Child Health Study Emotional Behavioural Scales (OCHS-EBS) checklist and Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID) by parent/caregiver and youth informants. The arrows on the left correspond to interfactor correlations, values for which appear at the bottom of Table 5.

required the square root of AVE values to be larger than the interfactor 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 interfactor 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 reinterviewed 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

Table 1. Sample Characteristics.

Characteristic	2014 OCHS Study Sample ^a	Retest Sample
Youth		
<i>n</i>	10,802	280
Age, mean (SD)	10.63 (4.07)	10.11 (4.16)
Male, %	51.3	49.3
Parent/caregiver		
<i>n</i>	6537	180
Age, mean (SD)	41.54 (7.20)	41.39 (6.87)
Male, %	11.8	16.7
Family		
<i>n</i>	6537	180
Household income, \$100,500 (\$162,600)		\$114,000 (\$94,400)
mean (SD)		
Single parent, %	20.6	17.2

OCHS, Ontario Child Health Study.

^aWeighted according to the probability of selection.

Table 2. Weighted Variability and Reliability of the OCHS-EBS by Gender and Age.^a

Disorder	Mean (SD)			Internal Consistency (α)	Test-retest reliability (r) ^b
	All	Male	Female		
Parent/caregiver Report (ages 4 to 11), $n = 6203$					$n = 148$
GAD	1.40 (2.03)	1.41 (2.04)	1.38 (2.02)	0.81	0.73
SAD	1.29 (2.04)	1.21 (1.99)	1.36 (2.08)	0.8	0.77
MDD	1.11 (1.70)	1.24 (1.84)	0.97 (1.52)	0.7	0.78
SP	1.61 (1.93)	1.53 (1.91)	1.70 (1.95)	0.81	0.7
ADHD	2.83 (3.21)	3.38 (3.43)	2.24 (2.83)	0.87	0.76
ODD	1.56 (1.98)	1.82 (2.16)	1.28 (1.73)	0.79	0.77
CD	0.33 (0.89)	0.43 (1.01)	0.22 (0.73)	0.62	0.71
Parent/caregiver report (ages 12 to 17), $n = 4292$					$n = 105$
GAD	1.89 (2.38)	1.75 (2.34)	2.04 (2.42)	0.85	0.84
SAD	0.72 (1.59)	0.65 (1.51)	0.81 (1.67)	0.81	0.79
MDD	1.87 (2.55)	1.72 (2.48)	2.03 (2.61)	0.83	0.75
SP	1.78 (2.25)	1.70 (2.30)	1.87 (2.20)	0.86	0.78
ADHD	2.28 (2.93)	2.71 (3.19)	1.81 (2.54)	0.87	0.87
ODD	1.68 (2.22)	1.80 (2.31)	1.54 (2.10)	0.84	0.7
CD	0.44 (1.36)	0.52 (1.55)	0.36 (1.11)	0.8	0.82
Youth report (ages 12 to 17), $n = 3925$					$n = 96$
GAD	3.26 (3.05)	2.67 (2.75)	3.90 (3.23)	0.86	0.78
SAD	1.95 (2.49)	1.64 (2.21)	2.29 (2.72)	0.79	0.54
MDD	2.86 (3.32)	2.30 (2.75)	3.46 (3.76)	0.85	0.78
SP	3.08 (2.69)	2.75 (2.57)	3.44 (2.78)	0.84	0.78
ADHD	3.79 (3.19)	3.85 (3.26)	3.71 (3.11)	0.81	0.74
ODD	2.07 (2.20)	2.05 (2.16)	2.08 (2.24)	0.76	0.82
CD	0.90 (1.50)	0.97 (1.58)	0.83 (1.42)	0.66	0.6

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; ODD, oppositional defiant disorder; SAD, separation anxiety disorder; SP, social anxiety disorder (social phobia).

^aThe 2014 Ontario Child Health Study (OCHS) sample was weighted based on the probability of selection. Retest subsample was unweighted.

^bAll estimates $P < 0.01$.

sampled evenly across the 3 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 1 or more different disorder scales, or high correlations with other items (results available in the 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 1 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 α for internal consistency, and test-retest reliabilities for caregiver report for ages 4 to 11 and 12 to 17, as well as youth report for ages 12 to 17. The scales comprise the same items across samples and informants. With the exception of youth-assessed CD and caregiver-assessed CD for ages 4 to 11, reliability estimates were all over 0.70 with 1 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 3 groups, respectively.

Internal convergent and discriminant validity. Table 3 summarizes the convergent and discriminant validity of the scales. Except for youth report CD, 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 of 42 comparisons in the caregiver model and 25 of 42 comparisons in the youth model. Discriminant validity test failures

Table 3. Weighted Scale AVE Values, Interfactor Correlations, and Fit Indices for Confirmatory Factor Analysis by Informant.

Characteristic	AVE ($\sqrt{\text{AVE}}$)	Interfactor 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	0.88
Model fit indices							
χ^2 (df) ^a	2264.367 (df = 1253)	P < 0.001					
CFI	0.943						
RMSEA	0.009						
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	0.84
Model fit indices							
χ^2 (df) ^a	3808.501 (df = 1253)	P < 0.001					
CFI	0.970						
RMSEA	0.014						

ADHD, attention-deficit hyperactivity disorder; AVE, average variance extracted; CD, conduct disorder; CFI, comparative fit index; GAD, generalized anxiety disorder; MDD, major depressive disorder; ODD, oppositional defiant disorder; RMSEA, root mean squared error of approximation; SAD, separation anxiety disorder; SP, social anxiety disorder (social phobia).

^a χ^2 test

Table 4. Multitrait, Multimethod Matrix Showing Point-Biserial Correlations between the OCHS-EBS Scale Scores and MINI-KID Disorder Classifications by Informant.^a

Trait	MINI-KID-P							MINI-KID-Y						
	GAD	SAD	MDD	SP	ADHD	ODD	CD	GAD	SAD	MDD	SP	ADHD	ODD	CD
OCHS-EBS-P														
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							
OCHS-EBS-Y														
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							

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-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.

^aAll correlations P < 0.01.

resulted from interfactor correlations being larger than the square root of AVE for GAD (MDD), MDD (GAD, ODD), ODD (CD, MDD), and 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 of 30 cases; test failures resulted from higher square root AVE values than interfactor correlations for GAD (MDD), MDD (GAD), ADHD (MDD, ODD), and ODD (MDD, ADHD).

Table 5. Weighted Standardized Factor Loadings, AVE Values, Interfactor Correlations, and Fit Indices for Confirmatory Factor Analysis—Parent/Caregiver and Youth Informant.

Model	Parent/Caregiver		Youth				
	Standardized Factor Loadings (Error Variance)	AVE ($\sqrt{\text{AVE}}$)	Standardized Factor Loadings (Error Variance)	AVE ($\sqrt{\text{AVE}}$)			
OCHS-Int		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)				
OCHS-Ext		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)				
MINI-Int		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)				
MINI-Ext		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)				
Interfactor Correlations ^a							
	P-OCHS-Int	P-OCHS-Ext	P-MINI-Int	P-MINI-Ext	Y-OCHS-Int	Y-OCHS-Ext	Y-MINI-Int
P-OCHS-Ext	0.73						
P-MINI-Int	0.76	0.48					
P-MINI-Ext	0.47	0.71	0.57				
Y-OCHS-Int	0.43	0.27	0.48	0.23			
Y-OCHS-Ext	0.43	0.54	0.45	0.49	0.75		
Y-MINI-Int	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
Model fit indices							
χ^2 (df) ^b	745.799 (df = 323)		P < 0.001				
CFI	0.953						
RMSEA	0.011						

ADHD, attention-deficit hyperactivity disorder; AVE, average variance extracted; CD, conduct disorder; CFI, comparative fit index; 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; RMSEA, root mean squared error of approximation; SAD, separation anxiety disorder; SP, social anxiety disorder (social phobia); Y, youth.

^aThese correlations correspond to the paths identified in Figure 1.

^b χ^2 test

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 caregivers 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 caregiver-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 intercorrelations 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 caregiver and youth instruments together with scoring instructions and are provided as online supplemental material.

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-reported (ages 4 to 11) CD while test-retest reliability was less than $r = 0.70$ for youth-reported CD. Although the internal convergent validity criterion was met for caregiver-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.⁴¹ 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, a number of internal discriminant validity failures occurring between disorders overlapped 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,^{38,40} 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 caregiver and youth informants was demonstrated for individual disorders in 81 of 84 comparisons. Again, exceptions occurred between disorders with similar characteristics (GAD and MDD as well as GAD and SP). When CFA was used to compare second-order factors representing externalizing and internalizing disorders derived for each informant (caregiver, 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 article 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 article 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 article.

Appendix

Table A1. 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
SAD	When anxious, his or her mind goes blank	.85
	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
MDD	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
	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

(continued)

Table A1. (continued)

Disorder	Item	Factor Loading
SP	Talks about killing self (youth item: thinks about killing self)	.96
	Doesn't like to be with people he or she doesn't know ^a	.83
	Afraid of doing things in front of others	.85
	Avoids social situations	.95
ADHD	Is nervous with people he or she doesn't know	.86
	Gets anxious about meeting new people	.88
	Makes careless mistakes	.69
	Can't concentrate, can't pay attention for long	.82
	Fails to finish things he or 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 ^a	.78
	Uses weapons when fighting	.81
	Has been physically cruel to others	.87
	Destroys things belonging to his or her family or other children	.72
	Has broken into someone else's house, building, or car	.87
	Sets fires ^a	.85
	Steals outside the home ^a	.81
	Stays out at night despite being told not to	.71
Runs away from home	.85	
Truancy; skips school	.70	

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).

^aItem did not meet rating criteria but added based on high clinician ranking indicating a core disorder symptom. Based on confirmatory factor analysis with 59 items, 7 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').

Data Access

Data access available through Statistics Canada Research Data Centres.


Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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 funding the Ontario Ministry of Children and Youth Services and the Ontario Ministry of Education. Dr. Boyle was supported by a 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 is supported 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.

ORCID iD

Laura Duncan, MA  <http://orcid.org/0000-0001-7120-6629>

Supplemental Material

Supplemental material for this article is available online.

References

- 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.
- Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington: Department of Psychiatry, University of Vermont; 1991.
- Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry.* 2001;40(11):1337-1345.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington (VA): American Psychiatric Publishing; 2013.
- 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.
- Kraemer HC, Noda A, O'Hara R. Categorical versus dimensional approaches to diagnosis: methodological challenges. *J Psychiatr Res.* 2004;38(1):17-25.
- 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.
- 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.
- 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(2):189-213.
15. Statistics Canada. 2014 Ontario Child Health Study (master file). Survey No. 3824. Ottawa (ON): Statistics Canada; 2017.
16. Government of Canada. Statistics Act. 2018 [cited 2017 June 7]. Available from: <http://laws-lois.justice.gc.ca/eng/acts/S-19/index.html>
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(suppl 20):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(3):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):304-321.
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, ed. *Comprehensive handbook of personality and psychopathology, adult psychopathology*. Vol. 2. New York (NY): John Wiley; 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;18(1):39-50.
34. Hair JF, Anderson RE, Babin BJ, et al. *Multivariate data analysis: a global perspective*. Vol. 7. Upper Saddle River (NJ): Pearson; 2010.
35. Brown TA. *Confirmatory factor analysis for applied research*. New York (NY): Guilford; 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):111-133.
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):121-153.
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.