Psychiatry Research 278 (2019) 235–241

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Scale development and psychometric properties of internalizing symptoms: The interRAI Child and Youth Mental Health internalizing subscale

Chloe Lau[a,*], Shannon L. Stewart[b], Donald H. Saklofske[a], John Hirdes[c]

[a] Faculty of Social Sciences, Western University, London, Canada
[b] Faculty of Education, Western University, London, Canada
[c] School of Public Health and Health Systems, University of Waterloo, Waterloo, Canada

ARTICLE INFO

Keywords:
InterRAI
Psychometric
Internalizing
Mental health
Children
Depression
Anxiety
Anhedonia

ABSTRACT

Internalizing difficulties in children and adolescents refer to inner-directed disordered mood and emotions. The purpose of this study was to develop and assess the reliability and validity of the internalizing subscale on the interRAI Child and Youth Mental Health (ChYMH) in assessing broadband internalizing mental health symptoms. Data were collected from clinically-referred children/youths 4 to 18 years of age between 2015 and 2016 (N = 2536). After initial items relevant to internalizing disturbances underwent expert content validation, exploratory first-order and bifactor factor analyses and multidimensional item response theory (MIRT) parameterizations were conducted to test the validity of the measurement model. The internalizing subscale demonstrated strong measurement properties for a three-factor structure (i.e., depression, anxiety, anhedonia) and MIRT analyses showed individual items had acceptable discrimination parameters across the latent continuum. A series of four competing models using confirmatory factor analyses were conducted in a separate sample of 1397 children/youth assessed in 2017 and the bifactor model showed superior fit compared to other models. Finally, concurrent validity of this measure was confirmed based on relationships with other established subscales from criterion measures. The interRAI ChYMH internalizing subscale could be used to provide useful information for triaging and prioritizing referrals for children/youths.

1. Introduction

Internalizing symptoms have been widely recognized in child psychiatry as negative affectivity with disordered mood as well as overcontrolled behaviors consistent with symptoms of depression and anxiety (Achenbach and McConaughy, 1992; Beitchman et al., 1992). Epidemiological research suggested that stability in developmental psychopathology can be accounted for by broad internalizing and externalizing domains (e.g., Mesman and Koot, 2000). As such, internalizing conditions are synonymous to "emotional" disorders as externalizing conditions are to "behavioral" disorders. These characteristics can be identified as symptoms, syndromes, or diagnoses in the child and adult psychopathology literature (Compas et al., 1993). Internalizing symptoms are among the most chronic forms of psychopathology affecting children and adolescents (Last et al., 1996; Twenge and Nolen-Hoeksema, 2002). Specifically, homotypic continuity is observed for internalizing mental health symptoms in children as young as 2 to 3 years of age, and the cumulative prevalence of psychiatric disorders by age 16 for a depression and anxiety diagnosis is 9.9% and 9.5%, respectively (Costello et al., 2003; Mesman and Koot, 2000). The chronicity in internalizing symptoms highlights the need for comprehensive assessments to track symptoms throughout the lifespan for early identification and intervention across multiple service sectors (Smith and Smith, 2010).

Aligned with this perspective, international assessments have been developed by interRAI, a non-profit international organization network of 50 members from over 30 countries who intended to create standardization of items across assessment tools in different sectors (Bernabei et al., 2008). The interRAI assessment instruments are used along the service continuum (e.g., mental health, homecare, palliative care, emergency department) designed to assess and monitor symptoms and treatment outcomes across age groups and vulnerable populations (Hirdes et al., 2002). There are more than 20 instruments built with a core set of items relevant for all healthcare sectors with identical definitions, time frame observation period, and method of scoring (Hirdes et al., 2011). This compatibility of assessment decreases assessment burden and outcome tracking for smooth transitioning across multiple treatment settings (Stewart et al., 2015). Reliability and

* Corresponding author at: Department of Psychology, Social Sciences Centre, Rm. 7313, Western University, Ontario, Canada N6A 5C2.
E-mail address: clau263@uwo.ca (C. Lau).

https://doi.org/10.1016/j.psychres.2019.06.013
Received 20 November 2018; Received in revised form 10 June 2019; Accepted 11 June 2019
Available online 13 June 2019
0165-1781/

validity studies have demonstrated strong psychometric properties, test-retest reliability, and criterion validity for interRAI instruments (Hirdes et al., 2008; Lau et al., 2018; Stewart and Hamza, 2017).

The interRAI Child and Youth Mental Health (ChYMH) was developed for children 4 to 18 years of age accessing mental health care and is currently being used in over 60 public mental health agencies (Stewart et al., 2015). The interRAI ChYMH conceptualizes mental health conditions using a dimensional and holistic approach, in which the clinician considers the child's broader individual and environmental context for a symptomatically coherent pattern. Assessment is conducted through communicating with the child/youth and primary caregiver, observation, communication with healthcare providers, and review of medical records (Stewart et al., 2015). The interRAI ChYMH has a shared set of items allowing informational transfer with other child/youth interRAI instruments, such as ChYMH-Education (ChYMH-EDU), ChYMH-Developmental Disabilities (ChYMH-DD), the Youth Justice Custodial Facilities instrument (YJCF), Homecare Pediatrics, and the adult version interRAI Mental Health (Hirdes et al., 2011). The use of standardized assessment across agencies promotes efficient and effective care-planning, as well as promoting streamlined access to services across service sectors for vulnerable populations (Stewart and Hamza, 2017).

1.1. The present study

With a large number of healthcare facilities promoting the usage of interRAI to enhance continuity of care, the development of a brief subscale using existing items on the interRAI ChYMH instrument to provide scale summary reports would suit the needs of a large network of Canadian mental health treatment facilities. Although the interRAI consists of many mental health indicator items, no subscale exists to quantitatively capture broadband internalizing distress. Moreover, such a subscale would provide precise measurement and empirically quantified information on internalizing mental health symptoms. Conceptualizing broadband psychopathology as a continuous symptom-level phenomenon allows for analysis in both shared and unique variance in the hierarchy (Lahey et al., 2004). As such, lead agencies using the interRAI ChYMH assessment requested the incorporation of an internalizing mental health symptoms measure into the instrument to assist in examining placement and treatment progress.

The current study utilizes pre-existing items on the interRAI ChYMH assessment instrument to develop a measure that captures a broad spectrum of internalizing mental health symptoms in children and adolescents. This study addresses the following objectives: (1) establishing content validity through rational expert judgments, (2) an analysis of the underlying factor structure of the measure, and (3) establishing criterion validity. The expert content data was collected to validate the scales in this study, but the dataset used for the latter steps were retrospective analyses of existing datasets collected by mental health agencies. Refer to Fig. 1 for a flow diagram of the study design.

2. Methods

2.1. Expert content validation

The interRAI ChYMH assessment consists of multiple sections intended to capture the child's psychological state, behaviours, cognitive functioning, independence in daily activities, physical health, stressors, and interpersonal environment. A subset of items developed as mental illness indicators on the interRAI ChYMH assessment were selected to undergo expert content validation. Experts were recruited on a voluntary basis to participate in the content validation process. An email was sent out to invited local mental health professionals in the area that linked each participant to an online survey (www.limesurvey.com) and participants were asked to indicate the extent to which each individual item was *representative* of the internalizing mental health construct. Experts assigned numerical ratings for the representativeness of each of the 22 items on a 4-point ordinal scale (1 = not representative, 2 = minimally representative, 3 = moderately representative, and 4 = strongly representative) for children (specified to be between the ages of 4 and 11), and adolescents (specified to be between the ages of 12 and 18) separately (Lynn, 1986). The items were presented in random order to avoid an ordering effect and the description provided for each item on the interRAI ChYMH assessment form was provided for the experts. The experts were provided a definition of internalizing symptoms involving symptoms of depressed mood, anxiety, and anhedonia.

Consensus estimates were calculated to determine the extent experts shared interpretation regarding whether these items represented internalizing symptoms (Hinkin et al., 1997; Polit and Beck, 2006). The content validity index (CVI; Waltz and Bausell, 1981), with expert ratings of relevance were calculated at the item level (I-CVI) and at the scale level (S-CVI; Polit et al., 2007). As recommended by Polit et al. (2007), the interpretation of k^* values were detailed in Fleiss (1981) and Cicchetti and Sparrow (1981) accounting for chance agreement, with values bounded between 0.40 to 0.59 rated as fair, 0.60 to 0.74 rated as good, and 0.75 to 1.00 rated as excellent. Items that received low modified kappa statistics (i.e., $k^* \leq 0.74$) were flagged unsuitable for this scale (Polit et al., 2007). S-CVI/UA is the proportion of items that are rated content valid (i.e., rating of 3 or 4) by all the experts, in which a lower limit of acceptability is set at 0.80 (Polit et al., 2007). An acceptable S-CVA/UA value would suggest the scale is representative of internalizing symptoms according to experts.

2.2. Construct validity of the internalizing subscale

Upon reviewing expert data, archival data were analyzed to determine whether the items of the internalizing subscale deemed content valid by the experts possessed desirable psychometric properties. Data collected between 2015 and 2016 using the interRAI ChYMH were complete for 2536 children/youth (M_{age} = 12.16, SD = 3.59; 57.6% males) across 34 sites in the Province of Ontario, Canada.

Each child/youth completed the full assessment of the interRAI ChYMH, a semi-structured interview assessing the child/youth's strengths, needs, functioning, and areas of risk to inform intervention for mental health needs. Assessors had at least 2 years of experience working in a mental health setting and were trained over a two full-day interRAI ChYMH training workshop. Every clinician was provided a users' manual for the interRAI ChYMH, which provided the intent, definition, suggested process to obtain the information, and coding for every item in the instrument (Stewart et al., 2015). The definition for each item is provided for universal interpretations of key terms used. The coding instructions provide standardized procedures for proper codes to score and report each of the responses. All ratings are required to be gathered within a three-day window (Stewart et al., 2015). Publications by the interRAI authors further detail the format and utility of ChYMH instrument (Stewart et al., 2015), and is also available from interRAI's website (The InterRAI Organization, 2017).

2.2.1. Unrestricted first-order and bifactor factor analysis

Bootstrap sampling (number of bootstrap samples = 500) was conducted to compute robust factor analysis. Parameter estimates were generated from bootstrap sampling of estimated asymptotic polychoric correlation matrix algorithm using Bayes modal estimation, as the method is robust to normality violations (Choi et al., 2011; Flora and Curran, 2004). The factor analysis was carried out by means of diagonally weighted least squares estimator and an oblique, Promin rotation was considered (Li, 2016; Lorenzo-Seva, 2013). An exploratory bifactor model is additionally performed for direct relationships between primary variables and higher order factors, which gives insight into theoretical constructs between latent constructs and observed variables (Reise, 2012).

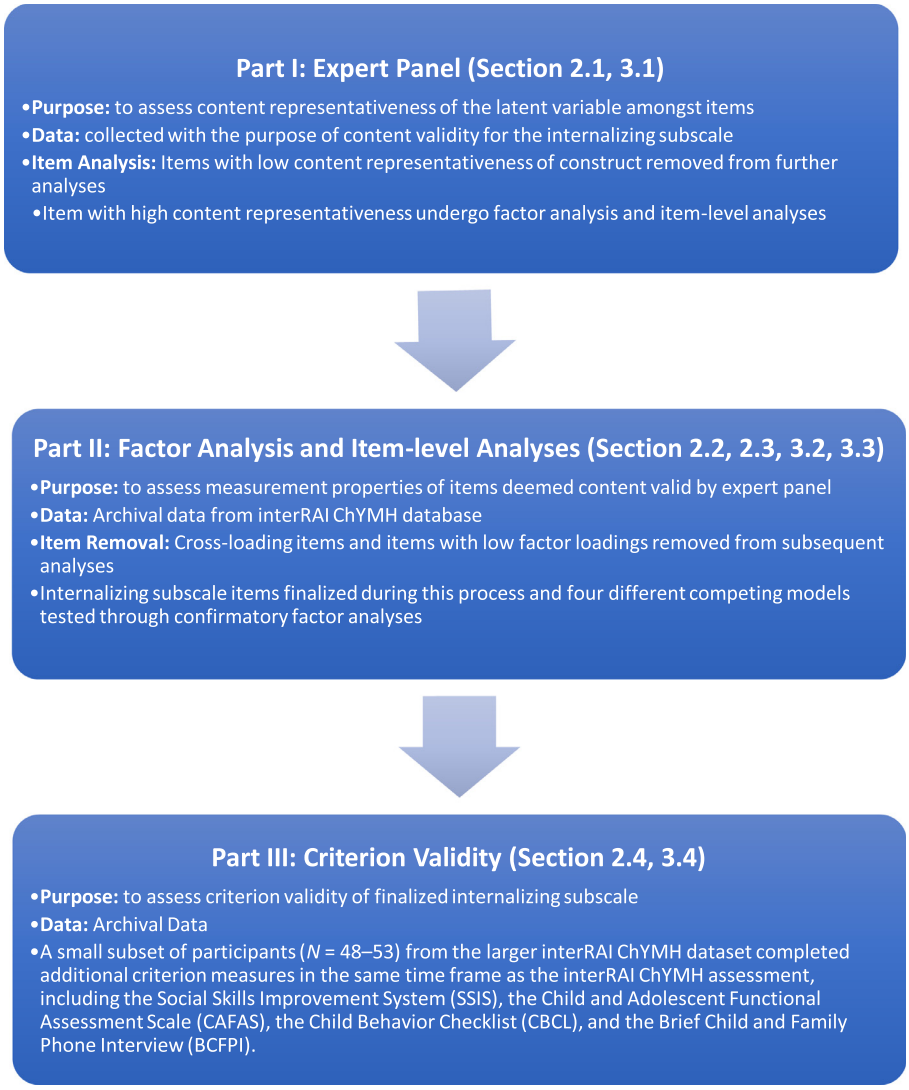

Fig. 1. Flow chart of overall study procedure.

2.2.2. Item response theory (IRT) parameterization

Reckase's (1985) multidimensional item response theory (MIRT) model was applied to assess item discrimination ability and category threshold. All factor analyses, reliability estimates, and item response theory parameterizations were conducted on the FACTOR software 10.5.02 for Windows 64-bits (Ferrando and Lorenzo-Seva, 2017).

2.3. Confirmatory factor analysis

To examine the reproducibility of the exploratory factor structure, data collected in 2017 from 1397 clinically referred children/youth (M_{age} = 12.44, SD = 3.64; 57.7% males) were examined. Four competing models using confirmatory factor analyses were conducted. Maximum likelihood estimation was used for parameter estimation (see Supplementary Materials A for details). As recommended by Byrne (2012) and Kline (2011), a root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) approximately 0.08 and 0.06 would suggest moderate and excellent model fit, respectively. A comparative fit index (CFI) and Tucker Lewis Index (TLI) in the range of 0.90 and 0.95 would suggest moderate and excellent model fit, respectively. Statistical analyses were conducted on JASP 0.8.5.1.

2.4. Establishing criterion validity

2.4.1. Participants

A small subset of participants (N = 48–53) independent from the interRAI ChYMH dataset completed additional criterion measures. Community-based data was collected from children/youth aged 6 to 18 years (M = 11.60, SD = 2.87; 75.6% males) were assessed using the interRAI ChYMH and additional criterion measures all conducted within a three-day window. Clinicians administering these measures and participants completing these measures were not aware of the present study hypotheses. The criterion measures were completed with the interRAI ChYMH assessment before the item "hypervigilance" was introduced into the assessment instrument, and hence, the item is missing for all participants in this smaller subset. The missing "hypervigilance" value was replaced with the mean of the sum of the remaining three variables in the anxiety factor (i.e., repetitive anxious complaints, unrealistic fears, episodes of panic) to compute total scores, which correlated highly with the original subscale in the larger dataset (r = 0.97, p < .001)

2.4.2. Measures

2.4.2.1. interRAI ChYMH internalizing subscale. The final interRAI ChYMH internalizing subscale consists of 12 items rated on a 5-point ordinal scale (0 = not present to 4 = exhibited daily in last 3 days, 3

or more episodes or continuously). Scores in this measure range from 0 to 48, with higher scores revealing greater frequency and severity of internalizing symptoms.

2.4.2.2. Social Skills Improvement System (SSIS). The SSIS is a parent or caregiver-report measure of social skills and problematic behaviours (Gresham et al., 2010). The SSIS internalizing scale consists of 7 items measuring emotional disturbance indicators (e.g., withdraws from others, acts anxious with others).

2.4.2.3. Child Behavior Checklist (CBCL). The CBCL is a set of standardized measures for children and adolescents between the ages of 4 to 18 years (Achenbach and Edelbrock, 1991). This measure consists of an internalizing broadband measures (i.e., anxious, depressive, and over-controlled) and specific narrowband subscales (i.e., social withdrawal, anxiety/depression, social problems, somatic complaints)

2.4.2.4. Child and Adolescent Functional Assessment Scale (CAFAS). The CAFAS is a clinician-administered measure that assesses functioning across several domains. The CAFAS moods/emotion subscale was used for this study and demonstrates reliability, concurrent validity, and discriminant validity (Hodges and Wong, 1996).

2.4.2.5. Brief Child and Family Phone Interview (BCFPI). The BCFPI is a parent- or caregiver-report standardized measure with 9 subscales embedded and specifically, the managing mood, managing mood and self-harm, and internalizing behaviours subscales were used for this study (Cunningham et al., 2009; Boyle et al., 2009). In the BCPFI subscales, the separation from parents, managing anxiety, and managing mood subscale create the internalizing subscale.

2.4.3. Statistical analysis

Bayesian correlations tests were conducted between interRAI internalizing subscale scores with various subscales on criterion measures. Jeffreys's Bayes Factor (1961) computes the probability of the observed data under the null hypothesis using a prior probability (i.e., probability hypothesis is true pre-data collection) and posterior probability (probability hypothesis is true post-data collection). Correlation tests proposed by Jeffreys (1961) were used, assuming bivariate normal distribution and a uniform, default prior on rho (Wagenmakers, 2007). Interpretation of Bayes Factors as evidence for alternative hypotheses with Bayes Factor of 1–3 as weak, 3–10 as substantial, 10–30 as strong, 30–100 as very strong, and >100 as decisive (Jarosz and Wiley, 2014). All Bayesian correlation analyses were conducted using JASP 0.8.1.1.

3. Results

3.1. Expert content validity

The final panel consisted of 15 participating experts (i.e., 4 males, 11 females; 3 in psychiatry, 11 in clinical psychology, 1 in psychiatric nursing) examining 22 items. Overall, 15 of 22 items demonstrated excellent content validity (I-CVI > 0.78 and $k^* > 0.74$) across both age groups. The 15 items deemed content valid (i.e., representative of internalizing symptoms in children and adolescents) during the expert panel evaluation process were selected to undergo unrestricted factor analysis.

3.2.1. Unrestricted factor analysis

Table 1 provides details for demographic information. The Hull Method with comparative fit index (CFI) deemed a three-factor solution had the best fit (Lorenzo-Seva et al., 2011). Upon removing three items with low discrimination values and cross-loadings, the remaining 12 items were subjected to unrestricted factor analysis using a polychoric correlation dispersion matrix, and the diagonally weighted least squares estimator. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.91 with a significant Bartlett's Statistic (χ^2 (66) = 9737.3, $p < .001$), indicating that the data were suitable for factor analysis.. The rotated factor loadings demonstrated that the items clustered to load onto three interpretable factors of anhedonia (α = 0.77), anxiety (α = 0.71), and depression (α = 0.82) dimensions. The first factor (i.e., anhedonia) explained 44.50% of the common variance between items with factor loadings ranging from 0.63 to 0.91. The second factor (i.e., anxiety) accounted for an additional 13.12% of variance with factor loadings ranging from 0.62 to 0.88 and the third factor (i.e., depression) accounted for an additional 11.74% of variance with factor loadings ranging from 0.66 to 0.93. The three-factor structure had factor loadings were in the acceptable range (>0.30) onto the primary factor and no significant cross-loadings were observed. The S-CVI/UA for the 12-item internalizing subscale is 0.89 and 0.97 for children (i.e., ages 4–11) and adolescents (i.e., ages 12–18), respectively, suggesting the scale has high representativeness of internalizing symptoms as rated by experts.

3.2.2. Exploratory bifactor model

Exploratory bifactor analysis was carried out by means of diagonally weighted least squares and the model specified included a three-factor structure along with a general factor (GF or “internalizing”) that loads directly onto the primary variables (Holzinger and Swineford, 1937). RMSR for this matrix is 0.02. Factor loadings ranged from 0.33 to 0.73 in an internalizing general factor (GF) and all factor loadings remained high in their respective three factors, with the exception of the item “decreased energy” which loaded strongly onto the internalizing GF (Table 2). These results show that there is the presence of a general factor (i.e., internalizing symptomology) as well as more specific factors (i.e., depression, anxiety, anhedonia) as part of the explanatory model (Reise et al., 2007).

3.2.3. Multidimensional item response theory

Item discrimination parameters for the anhedonia factor (a_1) were between 0.87 and 1.66. The item discrimination parameters for the anxiety factor (a_2) and depression factor (a_3) ranged from 0.85 to 1.53 and 1.01 to 2.01, respectively. Consistent with the item discrimination parameters (a_j) values, MDISC for items ranged from 0.85 to 2.04, suggesting acceptable discrimination values across individual facets and the GF. The thresholds for the lowest internalizing item (b_1) ranged from −0.46 to 1.06, while the threshold for the highest category (b_4) ranged from 1.37 to 3.34 on a *z*-score scale. These values suggest that items adequately capture symptomology of children/youth with high and low levels of internalizing symptoms (Table 3).

3.3. Confirmatory factor analysis

A series of competing models were conducted using confirmatory factor analyses to examine fit indices. Sample characteristics and further details on competing models were reported in Supplementary Materials A. Specifically, a one factor model, two factor model (i.e., anxiety and depression/anhedonia), three factor model (i.e., anhedonia, anxiety, depression), and a bifactor model (i.e., anhedonia, anxiety, depression, global internalizing) were compared. Overall, the bifactor model showed superior fit compared to other competing models, $\chi^2(36)$ = 89.80, $p < .05$, CFI = 0.998, TLI = 0.996, RMSEA = 0.03, SRMR = 0.03. Factor loadings ranged from 0.44 to 0.69 for anhedonia, 0.39 to 0.76 for anxiety, 0.37 to 0.85 for depression, and 0.30 to 0.89 for the internalizing global factor. These values suggest that the scale can be conceptualized as capturing internalizing symptoms as well as its subdomains (i.e., depression, anxiety, anhedonia).

Table 1
Demographics for Children/Youth Assessed using the interRAI ChYMH (N = 2535).

	Number (% of sample)	
Gender	Male	1461 (57.6%)
	Female	1074 (42.7%)
Patient type	Inpatient	160 (6.3%)
	Outpatient	2375 (93.7%)
Assessment method	Person	1873 (73.9%)
	Phone	661 (26.1%)
	Video	1 (0.0%)
Legal guardianship	Both parents	1391 (54.9%)
	Only mother	753 (29.7%)
	Only father	106 (4.2%)
	Neither parent but other relative(s) or non-relative(s)	133 (5.2%)
	Child Protection Agency (e.g., CAS)	134 (5.3%)
	Public guardian	3 (0.1%)
	Youth cares for self	15 (0.6%)
DSM-IV Diagnosis (Diagnosed/All Assessed)	Reactive attachment disorder	46/1894 (2.4%)
	Attention deficit Hyperactive disorder	1045/2019 (51.8%)
	Disruptive behaviour disorders (i.e., ODD, CD)	559/1959 (28.5%)
	Learning/Communication disorder	508/1940 (26.2%)
	Autism spectrum disorder	239/1922 (12.4%)
	Substance-related disorders	64/1926 (3.4%)
	Schizophrenia and other psychotic disorders	19/1932 (0.98%)
	Mood disorders	410/1885 (21.8%)
	Anxiety disorders	917/1920 (47.8%)
	Eating disorders	50/1923 (2.6%)
	Sleep disorders	74/1912 (3.8%)
	Adjustment disorders	72/1898 (3.8%)

3.4. Criterion validity

The interRAI ChYMH internalizing subscale correlated strongly with the CBCL internalizing scale (Pearson's ρ = 0.62), SSIS internalizing scale (Pearson's ρ = 0.61), the BCFPI internalizing scale (Pearson's ρ = 0.49). These are strong correlations categorized as “decisive” according to the corresponding Jeffreys's Bayes factor. Additional correlations between the internalizing subscale and criterion measure subscales with Jeffreys's Bayes Factor can be found in Table 4.

4. Discussion

Previous research has acknowledged the importance of measuring both broadband and narrow-band symptoms in childhood mental health (Achenbach et al., 2016; Forbes et al., 2016). The present findings follow the development and psychometric validation of the interRAI ChYMH internalizing subscale that showed promising psychometric properties and strong criterion validity. Following expert content validation, items that were deemed not representative of internalizing symptoms were flagged for deletion and the overall scale-content validity index deduced demonstrated that the final internalizing subscale utilized items that adequately captured the conceptualization and operationalization of internalizing mental health conditions (Hinkin et al., 1997). Item-level analyses and unrestricted factor analyses further reduced the number of cross-loading items with low discrimination values (i.e., provides noise in measurement). The resulting three-dimensional solution of anhedonia, anxiety, and depression demonstrated acceptable levels of internal consistency and construct validity.

Based on results from the factor analysis, the three-factor model was determined to be the most parsimonious, well-fitting model for the internalizing subscale. The internalizing subscale showed a dominant first factor of symptoms resembling anhedonia, which is the most specific symptom for young children with depression (Luby et al., 2003). Factor loadings were interpretable and no cross-loadings on the group factors were found, suggesting the bifactor and three-dimensional

Table 2
Exploratory bifactor solution of the 12-item internalizing scale.

Items	Anxiety (F1)	Depression (F2)	Anhedonia (F3)	Internalizing general factor
Repetitive anxious complaints/concerns	**0.556**	0.277	−0.027	**0.332**
Hypervigilance	**0.471**	0.144	−0.100	**0.490**
Unrealistic fears	**0.684**	0.203	−0.096	**0.408**
Episodes of panic	**0.476**	0.150	−0.041	**0.513**
Lack of motivation	−0.100	−0.071	**0.426**	**0.717**
Anhedonia	−0.032	0.187	**0.498**	**0.706**
Withdrawal from activities of interest	−0.029	0.060	**0.472**	**0.650**
Decreased energy	−0.061	−0.042	0.122	**0.725**
Made negative comments	−0.108	**0.694**	0.037	**0.509**
Self-deprecation	−0.010	**0.679**	−0.131	**0.548**
Expressions of Guilt/Shame	0.014	**0.389**	−0.274	**0.591**
Expressions of hopelessness	−0.107	**0.417**	−0.007	**0.668**
Interfactor correlations	**F1**	**F2**	**F3**	**General factor**
F1	1.000	–	–	–
F2	−0.029	1.000	–	–
F3	0.190	−0.032	1.000	–
General factor	0.000	0.000	0.000	1.000

Bold text represents factor loadings < 0.300.

Table 3
Reckase's multidimensional item response theory parameterization (1985) with item discrimination, MDISC, and category threshold values.

Items	Item discrimination dimension (a_1)	Item discrimination dimension (a_2)	Item discrimination dimension (a_3)	MDISC	Category threshold b_1	Category threshold b_2	Category threshold b_3	Category threshold b_4
Repetitive anxious complaints/ concerns	−0.111	**0.958**	0.069	**0.967**	−0.146	0.347	0.817	1.370
Hypervigilance	0.066	**0.850**	0.053	**0.854**	1.058	1.542	1.954	2.365
Unrealistic fears	−0.153	**1.527**	−0.071	**1.536**	0.361	1.135	1.687	2.342
Episodes of panic	0.169	**0.871**	0.010	**0.887**	0.240	1.195	1.897	2.521
Lack of motivation	**1.661**	−0.103	−0.177	**1.673**	0.378	1.086	1.694	2.153
Anhedonia	**1.632**	0.017	0.171	**1.641**	1.021	1.912	2.715	3.340
Withdrawal from activities of interest	**1.378**	0.021	−0.090	**1.381**	1.005	1.926	2.512	2.974
Decreased energy	**0.871**	0.025	0.130	**0.881**	0.210	0.698	1.293	1.859
Made negative comments	0.030	−0.185	**1.780**	**1.790**	−0.460	0.857	1.794	2.677
Self-deprecation	−0.288	0.105	**2.013**	**2.036**	−0.445	1.020	2.017	3.021
Expressions of guilt/shame	−0.100	0.153	**1.007**	**1.023**	0.320	1.202	1.893	2.527
Hopelessness	0.374	−0.091	**1.056**	**1.124**	0.416	1.440	2.166	2.842

Bold text represents discrimination values of items associated with the dimension.

model are equally interpretable (Reise et al., 2010). Finally, confirmatory factor analyses on a separate sample collected at a later date demonstrated the reproducibility of the bifactor model structure and this model had superior fit compared to competing models. These results suggest that clinicians using this measure could gather reliable and valid information from both the total score and subscale scores. Collectively, these results suggest clinicians may choose to use the internalizing global factor or the depression, anxiety, and anhedonia subdomains to derive meaningful information on the individual's symptomatology.

Multidimensional item response theory (MIRT) analyses further supported the validity of the measurement model, suggesting that items adequately reflect internalizing symptoms and capture children/youth with high and low levels of internalizing symptomology (Hambleton et al., 1991). In the present subscale, the item "made negative comments" is the least difficult item (i.e., children/youth assessed need less of the global internalizing trait to endorse this item) while the item "hypervigilance" is the most difficult item (i.e., children/youth assessed need more of the global internalizing trait to endorse this item). Hence, children/youth who endorse the item "hypervigilance" may have a higher probability of greater internalizing symptomology.

Finally, the interRAI ChYMH internalizing subscale was developed to assist with problem identification as well as monitor changes in clinical status over time. The strong criterion validity suggested that the internalizing subscale shows promising clinical utility. Almost all reported Bayes factors from the Bayesian correlations tests indicated that the data offered overwhelming support for the existence of the expected relationship between the interRAI ChYMH internalizing subscale and criterion measures. Further validation work is needed to examine the sensitivity of this scale to document changes in symptoms in individual patients over time.

Although this study used a large sample size of clinically referred youth across multiple mental health facilities across the province of Ontario, this study is not without limitations. The sample of clinically-referred youth were likely not representative of the child/youth population in general. In the absence of a normative sample, it remains unclear whether psychometric properties would replicate in children/youth without a wide range of physical and mental health problems. Moreover, the present study did not assess test-retest reliability, but provided that the current sample were assessed with interRAI as part of standardized assessment for treatment-planning during intake, a study investigating temporal stability would have to be conducted on separate samples who are not seeking treatment. Finally, this scale was designed to measure symptoms of depressed mood, anxiety, and anhedonia, and may not capture all symptoms related to internalizing distress (e.g., psychosomatic complaints). Future studies should examine how the internalizing subscale could be used concurrently with other subscales on the interRAI ChYMH assessment to create a comprehensive symptoms profile.

Overall, this study confirmed that the internalizing subscale shows strong reliability, structural validity, and criterion validity with other standard child and youth internalizing measures for detecting symptoms related to broadband emotional disturbances in children/youth. The relative economy of using subscales to detect frequency and severity of symptoms is beneficial to decreasing the assessment burden and involves less clinician time than a fully structured diagnostic interview during intake. The interRAI is copyrighted but usage is free of charge to researchers, clinicians, and government subject to the terms of a user license agreement with interRAI (www.interrai.org).

Declaration of Competing Interest

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

Table 4
Pearson's rho Bayesian correlations for ChYMH Internalizing scale and criterion measures of SSIS, CBCL, BCFPI, and CAFAS.

Criterion scale	Total Pearson correlation ρ (BF_{+0}(d))	Anhedonia subscale ρ (BF_{+0}(d))	Depression subcale ρ (BF_{+0}(d))	Anxiety subcale ρ (BF_{+0}(d))
SSIS: internalizing behaviour	0.61 (30,352)	0.492(177)	0.47 (82.56)	0.38 (7.63)
BCFPI: managing mood	0.50 (393)	0.631(44,600)	0.43 (26.92)	−0.06 (0.19)
BCFPI: managing mood and self-harm	0.55 (1396)	0.621(13,366)	0.50 (154)	−0.01 (0.18)
BCFPI: internalizing behaviours	0.49 (279)	0.46 (53.25)	0.40 (12.11)	0.10 (0.22)
BCFPI: Anxiety	0.23 (1.27)	0.265 (1.12)	0.13 (0.25)	0.07 (0.19)
BCFPI: social participation	0.47(199)	0.552 (2187)	0.33 (3.41)	0.04 (0.18)
CBCL: Internalizing	0.62 (15,310)	0.566 (767)	0.51 (130)	0.30 (1.48)
CBCL: social withdrawal	0.39 (16.13)	0.546 (528)	0.19 (0.40)	0.12 (0.24)
CBCL: anxiety/depression	0.61 (9574)	0.371(4.73)	0.63 (10,509)	0.37 (4.64)
CBCL: social problems	0.33 (4.50)	0.040 (0.19)	0.34 (2.70)	0.37 (4.84)
CBCL: somatic complaints	0.42 (6.24)	0.402(2.67)	0.36 (1.49)	0.15 (0.31)
CAFAS: mood/emotions	0.21 (1.18)	0.10 (0.21)	0.20 (0.51)	0.09 (0.20)

Note. BF_{+0}(d) = Jeffreys's Bayes Factor. Interpretation of Bayes Factors as evidence for alternative hypotheses with Bayes Factor of 1–3 as weak, 3–10 as substantial, 10–30 as strong, 30–100 as very strong and >100 as decisive (Jarosz and Wiley, 2014).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2019.06.013.

References

Achenbach, T.M., Edelbrock, C., 1991. Child Behavior Checklist. ASEBA Burlington (Vt), 7.

Achenbach, T.M., Ivanova, M.Y., Rescorla, L.A., Turner, L.V., Althoff, R.R., 2016. Internalizing/externalizing problems: review and recommendations for clinical and research applications. J. Am. Acad. Child Adolesc. Psychiatry 55 (8), 647–656.

Achenbach, T.M., McConaughy, S.H., 1992. Taxonomy of internalizing disorders of childhood and adolescence. In: Reynolds, W.M. (Ed.), Wiley Series on Personality Processes. Internalizing disorders in Children and Adolescents. John Wiley & Sons, Oxford, England, pp. 19–60.

Bernabei, R., Landi, F., Onder, G., Liperoti, R., Gambassi, G., 2008. Second and third generation assessment instruments: the birth of standardization in geriatric care. J. Gerontol. A Biol. Sci. Med. Sci. 63 (3), 308–313.

Beitchman, J.H., Inglis, A., Schachter, D., 1992. Child psychiatry and early Intervention: II. The internalizing disorders. Can. J. Psychiatry 37 (4), 234–239. https://doi.org/10.1177/070674379203700405.

Boyle, M.H., Cunningham, C.E., Georgiades, K., Cullen, J., Racine, Y., Pettingill, P., 2009. The Brief Child and Family Phone Interview (BCFPI): 2. Usefulness in screening for child and adolescent psychopathology. J. Child Psychol. Psychiatry 50 (4), 424–431.

Byrne, B.M., 2012. A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models. Springer Science & Business Media.

Choi, J., Kim, S., Chen, J., Dannels, S., 2011. A comparison of maximum likelihood and Bayesian estimation for polychoric correlation using Monte Carlo simulation. J. Educ. Behav. Stat. 36 (4), 523–549.

Cicchetti, D.V., Sparrow, S.S., 1981. Developing criteria for the rating of specific items in a given inventory. Am. J. Ment. Defic. 86, 127–137.

Compas, B.E., Ey, S., Grant, K.E., 1993. Taxonomy, assessment, and diagnosis of depression during adolescence. Psychol. Bull. 114 (2), 323–344.

Costello, E.J., Mustillo, S., Erkanli, A., Keeler, G., Angold, A., 2003. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch. Gen. Psychiatry 60 (8), 837–844.

Cunningham, C.E., Boyle, M.H., Hong, S., Pettingill, P., Bohaychuk, D., 2009. The Brief Child and Family Phone Interview (BCFPI): 1. Rationale, development, and description of a computerized children's mental health intake and outcome assessment tool. J. Child Psychol. Psychiatry 50 (4), 416–423.

Ferrando, P.J., Lorenzo-Seva, U., 2017. Program FACTOR at 10: origins, development and future directions. Psicothema 29 (2), 236–240.

Fleiss, J.L., 1981. Balanced incomplete block designs for inter-rater reliability studies. Appl. Psychol. Meas. 5 (1), 105–112.

Flora, D.B., Curran, P.J., 2004. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol. Methods 9 (4), 466.

Forbes, M.K., Tackett, J.L., Markon, K.E., Krueger, R.F., 2016. Beyond comorbidity: toward a dimensional and hierarchical approach to understanding psychopathology across the life span. Dev. Psychopathol. 28 (4pt1), 971–986.

Gresham, F.M., Elliott, S.N., Cook, C.R., Vance, M.J., Kettler, R., 2010. Cross-informant agreement for ratings for social skill and problem behavior ratings: an investigation of the Social Skills Improvement System—rating scales. Psychol. Assess. 22 (1), 157–166.

Hambleton, RK., Swaminathan, H., Rogers, WH., 1991. Fundamentals of Item Response Theory. Sage Publications, Newbury Park.

Hinkin, T.R., Tracey, J.B., Enz, C.A., 1997. Scale construction: developing reliable and valid measurement instruments. J. Hosp. Tour. Res. 21 (1), 100–120.

Hirdes, J.P., Smith, T.F., Rabinowitz, T., Yamauchi, K., Pérez, E., Telegdi, N.C., ..., Fries, B.E., 2002. The Resident Assessment Instrument-Mental Health (RAI-MH): inter-rater reliability and convergent validity. J. Behav. Health Serv. Res. 29 (4), 419–432.

Hirdes, J., Curtin-Telegdi, N., Mathias, K., Perlman, C.M., Saarela, T., Kolbeinsson, H., ... & Martin, L., 2011. InterRAI mental health clinical assessment protocols (CAPs) for use with community and hospital-based mental health assessment instruments. Version 9.1.

Hirdes, J.P., Ljunggren, G., Morris, J.N., Frijters, D.H., Soveri, H.F., Gray, L., ..., Gilgen, R., 2008. Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. BMC Health Serv. Res. 8 (1), 277.

Hodges, K., Wong, M.M., 1996. Psychometric characteristics of a multidimensional measure to assess impairment: the Child and Adolescent Functional Assessment Scale. J. Child Fam. Stud. 5 (4), 445–467.

Holzinger, K.J., Swineford, F., 1937. The bi-factor method. Psychometrika 2 (1), 41–54.

Jeffreys, H., 1961. Theory of probability, (3rd ed.). Oxford University Press, Clarendon Press, Oxford.

Jarosz, A.F., Wiley, J., 2014. What are the odds? A practical guide to computing and reporting Bayes factors. J. Probl. Solving 7 (1), 2.

Kline, R.B., 2011. Principles and Practice of Structural Equation Modeling. Guilford Press, New York, NY.

Lahey, B.B., Applegate, B., Waldman, I.D., Loft, J.D., Hankin, B.L., Rick, J.., 2004. The structure of child and adolescent psychopathology: generating new hypotheses. J. Abnorm. Psychol. 113, 358–385.

Lau, C., Stewart, S.L., Saklofske, D.H., Tremblay, P.F., Hirdes, J., 2018. Psychometric evaluation of the interRAI Child and Youth Mental Health Disruptive/Aggression Behaviour Scale (DABS) and hyperactive/distraction scale (HDS). Child Psychiatry Hum. Dev. 49 (2), 279–289.

Last, C.G., Perrin, S., Hersen, M., Kazdin, A.E., 1996. A prospective study of childhood anxiety disorders. J. Am. Acad. Child Adolesc. Psychiatry. 35 (11), 1502–1510.

Li, C.H., 2016. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav. Res. Methods 48 (3), 936–949.

Lorenzo-Seva, U., Timmerman, M.E., Kiers, H.A., 2011. The Hull method for selecting the number of common factors. Multivar. Behav. Res. 46 (2), 340–364.

Lorenzo-Seva, U., 2013. Why Rotate My Data Using Promin? Department of Psychology, Universitat Rovira i Virgili, Tarragona Technical Report.

Luby, J.L., Heffelfinger, A.K., Mrakotsky, C., Brown, K.M., Hessler, M.J., Wallis, J.M., Spitznagel, E.L., 2003. The clinical picture of depression in preschool children. J. Am. Acad. Child Adolesc. Psychiatry 42 (3), 340–348.

Lynn, M.R., 1986. Determination and quantification of content validity. Nurs. Res. 35 (6), 382–386.

Mesman, J., Koot, H.M., 2000. Child-reported depression and anxiety in preadolescence: I. Associations with parent-and teacher-reported problems. J. Am. Acad. Child Adolesc. Psychiatry 39 (11), 1371–1378.

Polit, D.F., Beck, C.T., 2006. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res. Nurs. Health 29 (5), 489–497.

Polit, D.F., Beck, C.T., Owen, S.V., 2007. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res. Nurs. Health 30 (4), 459–467.

Reckase, M.D., 1985. The difficulty of test items that measure more than one ability. Appl. Psychol. Meas. 9 (4), 401–412.

Reise, S.P., 2012. The rediscovery of bifactor measurement models. Multivar. Behav. Res. Methods 47 (5), 667–696.

Reise, S.P., Morizot, J., Hays, R.D., 2007. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual. Life Res. 16 (1), 19–31.

Stewart, S., Hirdes, J., Curtin-Telegdi, N., Perlman, C.M., McKnight, M., MacLeod, K., ... & Berg, K., 2015. interRAI Child and Youth Mental Health (ChYMH) assessment form and user's manual: fFor use with in-patient and community-based assessments.

Stewart, S.L., Hamza, C.A., 2017. The Child and Youth Mental Health Assessment (ChYMH): an examination of the psychometric properties of an integrated assessment developed for clinically referred children and youth. BMC Health Serv. Res. 17 (1), 82.

Smith, J.P., Smith, G.C., 2010. Long-term economic costs of psychological problems during childhood. Soc. Sci. Med. 71 (1), 110–115.

The InterRAI Organization: Who we Areare. interRAI. 2017. Web site. http://www.interrai.org/organization.html. Accessed 30 January 2017.

Twenge, J.M., Nolen-Hoeksema, S., 2002. Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: a meta-analysis. J. Abnorm. Psychol. 111 (4), 578–588.

Wagenmakers, E.J., 2007. A practical solution to the pervasive problems of p values. Psychol. Bull. Rev. 14 (5), 779–804.

Waltz, C.F., Bausell, B.R., 1981. Nursing Research: Design Statistics and Computer Analysis. Davis FA.