Confirmatory factor structure of anxiety and depression: evidence of item variance across childhood

KHRISTA R. BOYLAN,¹ JESSIE L. MILLER,² TRACY VAILLANCOURT^{2,3} & PETER SZATMARI¹

1 Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, Ontario, Canada

2 Department of Psychology, Neuroscience and Behavior, McMaster University, Hamilton, Ontario, Canada

3 Faculty of Education and School of Psychology, University of Ottawa, Ottawa, Ontario, Canada

Key words

Anxiety, depression, oppositional, longitudinal, sex differences

Correspondence

Khrista Boylan, Department of Psychiatry and Behavioral Neurosciences, McMaster University, 1200 Main St West, Box 2000, Hamilton, Ontario, L8N 3Z5, Canada.

Telephone (+1) (905) 521–2100, ext 77250

Fax (+1) (905) 521-7938

Email: boylank@mcmaster.ca

Received 4 December 2009;

revised 4 June 2010;

accepted 18 November 2010

Abstract

The distinctiveness of anxiety and depressive symptoms in children has previously been questioned based on their high degree of comorbidity, shared risk factors, and treatment response. Developing children may show an unstable presentation of anxiety and depressive symptoms that would complicate interpretation of studies of comorbidity. The present study examined the measurement stability of anxiety and depressive symptoms across time and sex using a large epidemiologic sample of children. A nationally representative cohort of 1329 children (624 girls and 705 boys) aged four to seven in 1994 were drawn from the National Longitudinal Survey of Youth (NLSY). Using eight years of prospective data we examined whether a one or two factor structure of anxiety (five items) and depressive (four items) symptoms would be invariant across time and sex. Despite item variability within each factor across time, confirmatory factor analysis revealed distinct factors for anxiety and depression that were stable across time and sex. Results provide support that covariation between anxiety and depression is not likely the result of measurement overlap. However, items indicating factors of anxiety and depression in the NLSY may not be sufficient to permit developmentally-sensitive measurement of these factors. Copyright © 2011 John Wiley & Sons, Ltd.

Introduction

Anxiety and depression are among the two most prevalent psychiatric illnesses that occur during childhood and adolescence. Epidemiological studies indicate that 10–15% of children under the age of 12 meet criteria for an anxiety disorder and 3–5% with a depressive disorder (Axelson and Birmaher, 2001).

Anxiety and depression are defined as distinct syndromes among children, adolescents and adults (APA, 2001) and yet 33–75% of children meeting diagnostic criteria for major

depressive disorder have one or more comorbid anxiety disorders (Angold *et al.*, 1999). Conversely, comparatively fewer (5–69%) children classified as having an anxiety disorder had a comorbid depressive disorder (Angold *et al.*, 1999). The combined odds ratio of association between these two conditions ranged from five to 33 across 21 cross-sectional studies reviewed by Angold *et al.* (1999), suggesting that each condition is more likely to occur in the presence of the other than chance alone would predict. These observations may indicate that there are shared biological or contextual processes contributing to the co-occurrence of anxiety and depression.

Alternatively, rates of co-occurrence may vary because of the way anxiety or depression is measured. If similar items contribute to both the measurement of anxiety and depression then these disorders may be more likely to co-occur because of measurement overlap rather than because of true comorbidity. Age of the child and whether comorbidity is assessed cross-sectionally or longitudinally at the time of measurement may also be relevant, particularly if there is a progression between the disorders that is developmentally sensitive.

Measuring anxiety and depression symptoms across time

Over the past two decades, studies examining whether clusters of anxiety and depression symptoms are distinct latent constructs have predominantly used the original and revised versions of the "Tripartite Model" of anxiety and depression as a baseline model (Clark and Watson, 1991; Mineka *et al.*, 1998; Anderson and Hope, 2007). In the earlier model, anxiety and depression are composed of a triad of symptoms consisting of physiological hyperarousal (tremulousness, trouble breathing), low positive affect (anhedonia or hopelessness) and high negative affect (sad or irritable mood, worry). The premise is that while physiological hyperarousal is a collection of symptoms specific to anxiety, low positive affect is specific to depression and negative affect is common to both. A more recent revision of the model (Chorpita *et al.*, 2000; Laurent *et al.*, 2004) attempts to take into account the heterogeneity (i.e. fear versus worry items) amongst anxiety disorder symptoms. However, studies that do not separate the fear from worry items demonstrate support for a one-factor (Cole *et al.*, 1997) or two-factor (Cole *et al.*, 1997; Duffy *et al.*, 2005; Huebner and Dew, 1995; Joiner *et al.*, 1996; Laurent and Ettelson, 2001; Lee and Reebok, 2002; Lonigan *et al.*, 2003) model of anxiety and depression. This is important as it suggests that the method of assessment of anxiety and depression will affect their joint measurement structure.

Furthermore, key questions remain regarding the stability of factors of anxiety and depression across different ages of children, and also between boys and girls. This is of particular importance in the measurement of anxiety and depression, as it is known that temporally, anxiety precedes depression in most cases (Cicchetti and Toth, 1998; Kovacs *et al.*, 1989). Only two published studies have assessed the stability and uniqueness of anxiety and depression over time using a longitudinal design. Cole *et al.* (1998) followed a cohort of 330 nine-year-old children over three years and found that self or parent reported anxiety and depression measured using the Revised

Children's Manifest Anxiety Scale (RCMAS) (Reynolds, 1980) and the Children's Depression Inventory (CDI) (Kovacs, 1985) were independent and time-invariant constructs. This study did not formally test the independence of factor structure of the RCMAS and the CDI, but tested whether their path model was significantly different when overlapping items from both scales were removed. Lonigan *et al.* (2003) followed a cohort of 270 youth aged 9–16 years and found that the correlations between the RCMAS, the CDI and the Positive and Negative Affect Scale (Watson and Clark, 1988) were invariant across two occasions of measurement separated by seven months. Thus, in both of these longitudinal studies, the independence of depression and anxiety constructs were not formally tested, but inferred using other methods.

The current study was designed to specifically investigate whether the measurement structure of anxiety and depressive symptom factors conforms to a one- or two-factor model and whether this factor structure holds in a much larger sample ($N = 1329$) of children followed prospectively for a much longer period of time (eight years) than the Cole *et al.* (1998) and Lonigan *et al.* (2003) studies. Examining a period of eight years allowed us to test the stability of anxiety and depression across important socio-emotional and biological milestones. Another important feature of the study is the assessment of the stability of these constructs using a nationally representative sample of children. Such an approach provides robust evidence regarding the longitudinal association of anxiety and depression, independent of issues of sampling bias and severity of illness. In addition, it is important for epidemiologic researchers to have supporting evidence that brief measures of anxiety and depression that are commonly used in large epidemiologic surveys like the National Longitudinal Survey of Youth (NLSY) are invariant over time by age and sex. In such studies, brief measures with small item pools may not permit sufficient differentiation of anxiety from depression. Such knowledge is key in interpreting conclusions drawn regarding developmental differences if measurement invariance of items has *previously been demonstrated*.

Age and sex differences

The prevalence rates of depression and anxiety have been shown to differ by both age and sex (Angold *et al.*, 1999). For girls, depressive symptoms increase significantly at age 12 and the rate of Major Depressive Disorder (MDD) diagnoses double at that time (Hankin *et al.*, 1998). However prior to this age, rates of MDD do not differ by sex, with an overall prevalence rate of childhood MDD estimated at 2–8%.

For anxiety disorders in childhood the 12 month prevalence rates are estimated to be in the range of 10 to 20% (Costello *et al.*, 2004), and they occur more frequently in girls (ratio of 2:1). One large community study found a two-fold increase in the incidence of anxiety disorders among females, beginning around age six (Lewinsohn *et al.*, 1998). These sex and age differences in anxiety and depression appear to be robust, however, given that there have been only two longitudinal studies of the factor structure of anxiety and depression to-date, it remains a possibility that the purported age and sex differences reflect a degree of measurement error rather than true developmental changes, particularly if the items are not sensitive to developmental changes in the presentation of anxiety and depression in children.

Considering these findings, the purpose of the present study was to examine the stability of the factor structure of anxiety and depression across sex and across important developmental periods in a large nationally representative sample of American children. Our primary hypothesis was that anxiety and depression would be represented as two stable factors invariant of time and sex.

Method

Participants

Data were drawn from the NLSY 1979 – Child and Youth sample, a nationally representative sample of the offspring of 12,686 American women who were all between 14 and 21 years of age by December 31, 1978. Methods used in the study are explained in the NLSY User's Guide (Center for Human Resource Research, 2002). Informed consent was obtained from these women in compliance with the policies of the US Bureau of Labor and Statistics. Assessments of each child as well as additional demographic and developmental information were collected from the mother. At the time of the child's birth, 57% of mothers were married and their average age was 25.8 [standard deviation (SD) = 2.2] and 65.2% of women were living with the father of the child at the time of interview (1994). Forty-nine percent of the mothers were non-Black non-Hispanic, 22% were Hispanic and 29% were Black. The average family income in 1994 was \$43,309 (SD = 21,212) and on average mothers had completed 12.6 years of formal schooling (SD = 2.3 years). Eight percent had a clinically significant score (>11) on the Center for Epidemiologic Studies Depression Scale (Radloff, 1997).

The sample cohort included all the children in the original NLSY sample (624 girls and 705 boys) who were aged four to six in 1994. This sampling year and age range

was chosen to permit biannual assessment of a cohort where ages would not overlap, the same measure – the Behavior Problems Index (BPI) – was used each year and the data were as recent as possible. Three time points within the longitudinal analysis of this cohort (years 1994, 1998, and 2002, or Times 1, 2 and 3) allowed for the temporal assessment of symptom stability across ages 4–14 years. The age range includes early childhood, middle childhood, and early adolescence allowing us to compare scores and factor structure across three developmental time points.

Missing data

A total of 83.8% of girls and 82.4% of boys had complete data at all three time points and 99.2% of girls and 92.9% of boys had complete data at Time 1 and in either of Time 2 or Time 3. Those with full and partial data did not differ on any of the following socio-demographic variables: maternal depression score (4.49 versus 4.41, $t(900) = 0.17$, $p = 0.87$), maternal education (12.5 versus 12.4 years, $t(901) = 0.68$, $p = 0.50$), maternal age at birth (26 versus 26 years, $t(901) = 0.08$, $p = 0.93$), family income (\$42,701.78 versus \$43,257.27, $t(580) = -0.11$, $p = 0.91$), single parent status (38% versus 33%, $\chi^2(1) = 1.95$, $p = 0.16$). Missing data were imputed using the Full Information Maximum Likelihood (FIML) imputation in Amos 6 (Arbuckle, 2006). The data imputed by FIML have been shown to have the least bias, particularly when there is a small amount (<20%) of missing data and the distribution of the variables are multivariate normal (Arbuckle, 1996; Little and Rubin, 1989; Shafer and Graham, 2002). Missing data was imputed within individuals at the item level yielding complete data on all 1329 participants.

Instruments and procedure

Mothers of children aged four to six at Time 1 were interviewed biannually about their child's problem behaviors in addition to other developmental, familial and social factors. Questions were derived from the BPI; a scale created specifically for the NLSY (Peterson and Zill, 1986). The BPI is based on the Child Behavior Checklist (Achenbach and Edelbrock, 1981) and as such, its items are suited for behavioral reports as the parent is not privy to the inner experience of the child. The full BPI consists of 28 items, and five sub-factors have been previously identified: aggression (internal consistency $\alpha = 0.80$), peer problems ($\alpha = 0.80$), hyperactivity ($\alpha = 0.80$), depression ($\alpha = 0.73$), and dependent ($\alpha = 0.64$). BPI items have discriminated between clinic and community children in the National Child Health Supplement (Peterson and Zill, 1986), in a

sample of six to eight year olds in the NLSY (Colder *et al.*, 2002) and between children from high and low conflict marriages (Zill and Snyder, 1981).

Specific BPI items ($n=9$) to represent anxiety and depression constructs were chosen based on conceptual overlap with core symptoms of the actual psychiatric disorders (APA, 2001). Specifically, we chose depression items to reflect anhedonia, as opposed to worry or negativity, anxiety items to reflect the cognitive and behavioral aspects of fear. Ten independent raters (four psychologists and six psychiatrists) with expertise in child psychopathology sorted items into anxious and depressed categories as a test of face validity. We then tested the internal consistency of these derived scales (reported later) and in each case there was one factor and no items were judged to require removal based on change in internal consistency.

The depression subscale included the four items: (1) child feels worthless or inferior, (2) child is unhappy/sad/depressed, (3) child is withdrawn, and (4) child complains no one loves them. The anxiety subscale included the five items: (1) child is high-strung/tense/nervous, (2) child is too fearful or anxious, (3) difficulty keeping their mind off thoughts, (4) child is too dependent on others, and (5) child worries too much. Parents reported on the frequency their child engaged in each of these behaviors during the past week using a three-point scale (one representing never to three equaling often) with scale scores ranging 1–12 for the depressive symptoms subscale and 1–15 for the anxiety symptoms subscale. Internal consistencies were moderate, ranging from $\alpha = 0.67$ (age 4–6) to $\alpha = 0.77$ (age 12–14) for the depression subscale, from $\alpha = 0.62$ (age 4–6) to $\alpha = 0.74$ (age 12–14) for the anxiety subscale.

Confirmatory factor analysis (CFA)

Longitudinal confirmatory factor analysis (CFA) conducted using Amos 7.0 (Arbuckle, 2006) was employed to address the study objectives. We used a multi-group modeling procedure to test for both item and configural invariance in our latent factors of anxiety and depression across time (three waves) and sex (male and female). We tested for configural invariance using the total sample ($N=1329$) by specifying a two-factor model of depression (four-indicator latent factor) and anxiety (five-indicator latent factor) for each of the three time periods. Thus, a six-group model was specified with a total of six latent variables (27 indicators). We did not specify a three-factor latent model (i.e. Tripartite Model) to these data as items specific to physiologic hyperarousal were not available. We tested the two-factor model of anxiety and depression against an alternative one-factor model to evaluate the possibility that anxiety and depressive symptoms are

best represented by one underlying latent dimension. We allowed the residual variances of the same indicator to correlate across time points, and the latent factors anxiety and depression were allowed to correlate concurrently. No equality or cross-group constraints were imposed on the initial baseline model so as to freely estimate the factor loadings, residual variances and correlations between the two latent factors. Next we tested for item invariance both across time and sex following the method described by Byrne (2004).

We used several practical fit indices to test for item and configural invariance as recommended by Jaccard and Wan (1995) and Miles and Shevlin (2007). These included the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the root mean square error of approximation (RMSEA) and Akaike's Information Criterion (AIC) and the chi square difference test for nested models. These fit indices were chosen because they each involve different statistical strategies for testing fit and when used to complement each other, provide a more conservative evaluation of model fit. The CFI and TLI measure how much variability exists between the hypothesized and the observed model, with scores >0.9 being considered evidence of acceptable fit of the data to the hypothesized model. The numerical value of any single fit index cannot be used as a basis for accepting or rejecting a model, but should be used in conjunction with other indicators of model fit. Other aspects to consider include an assessment of the potential influence of reliability of measures (as unreliability can increase the likelihood of not rejecting a dissimilar model), model complexity and parsimony, and factor loadings (Miles and Shevlin, 2007).

Results

Descriptive statistics

Table 1 presents the means, standard deviations, Cronbach's alphas and inter-correlations of anxiety and depression at Time 1 (age 6–8), Time 2 (age 9–11), and Time 3 (age 12–14) for the total sample ($N=1329$). Item total correlations are all above 0.3, the recommended factor loading per item (Nunnally and Bernstein, 1994).

Wave-to-wave correlations of the anxiety factors for the total sample are presented in Table 1. Significant differences were found for paired correlations between Times 1 and 3 and between Times 2 and 3 only ($t(1327) = -2.57$, $p=0.05$). For depression, correlations between Times 1 and 2 and Times 1 and 3 were significantly different from each other ($t(1327) = -2.56$, $p=0.05$). Concurrent correlations between anxiety and depression factors were the highest at Times 2 and 3 and were both significantly greater than Time 1 ($t(1327) = -2.03$, $p=0.02$).

Table 1. Wave-to-wave non-parametric correlations among anxiety and depression constructs across three age groups for the total sample ($N = 1329$)

	1	2	3	4	5	6
1. Anxiety Time 1	1.0	—	—	—	—	—
2. Anxiety Time 2	0.46	1.0	—	—	—	—
3. Anxiety Time 3	0.35	0.43	1.0	—	—	—
4. Depression Time 1	0.49	0.30	0.26	1.0	—	—
5. Depression Time 2	0.34	0.60	0.31	0.44	1.0	—
6. Depression Time 3	0.18	0.25	0.59	0.26	0.33	1.0
Means	6.06	6.38	6.35	4.51	4.74	4.73
Standard deviation(SD)	1.42	1.73	1.67	1.00	1.23	1.27
International consistency	0.62	0.70	0.74	0.67	0.71	0.77

Time 1, 1994; Time 2, 1998; Time 3, 2002. There is a four-year interval between each time measurement. All correlations are significant at $p < 0.001$.

Mean levels of anxiety and depression were statistically significantly different across time (for depression, $F(1327) = 16.625$, $p < 0.001$; for anxiety, $F(1327) = 17.541$, $p < 0.001$) and *post hoc* paired *t*-tests between consecutive time points showed that the mean score at Time 1 was statistically significantly lower than the mean score at either Time 2 or Time 3 which did not differ from each other (data not shown). Only level of anxiety at Time 1 differed by sex – higher in boys than girls ($t(1327) = 2.04$, $p < 0.05$).

Model estimation – factorial measurement structure

Our initial two-factor unconstrained model (Model 1) showed acceptable fit to the data (see Table 2). When the concurrent latent correlations between anxiety and depression were each fixed to 1.0, to specify a single latent factor (Model 2), this model had a less acceptable fit to the previous two-factor model on all indices, suggesting that a two-factor structure was a better fit to the data.

We noted that some items had low factor loadings, thus to determine whether these items might cross-load on the

opposite factor, modification indices were reviewed. All of these indices were significantly smaller than the threshold suggested by Bentler (2000), thus different factor-item assignments would not improve the model fit.

Next we tested whether the two-factor structure would be invariant across time (Model 3). Compared with the previous two-factor model without constraints, there was a significant difference according to the chi-square difference test, whereas no suspected loss according to practical model fit indices was observed (Table 2). Factor loadings did not change from the unconstrained model (Model 1) with standardized regression weights ranging 0.48–0.73 for depression and 0.48–0.67 for anxiety items. We examined the pattern of factor loadings as modification indices were not contributory, and noted poor factor loadings for the depression items “no love” and “withdrawn” and the anxiety items “too fearful” and “depends too much on others”. We allowed these parameters to be estimated freely across time and this significantly improved the model fit such that the model was not significantly different than the unconstrained two-factor model overall, but

Table 2. Summary of consecutive confirmatory factor model tests

Nested model step	χ^2 (DF)	CFI	RMSEA 95% CI	AIC	TLI
Model 1: Unconstrained two-factor model	632.40(282)	0.962	0.030 0.027–0.034	878.40	0.951
Model 2: Unconstrained one-factor model	1035.69(285)	0.922	0.045 0.042–0.047	1275.698	0.904
Model 3: two factor –invariance across time	654.5 (292)	0.960	0.032 0.029–0.035	900.491	0.952
Model 4: two factor –invariance across time and sex	1023.2 (598)	0.949	0.025 0.022–0.27	1516.42	0.937

Note: χ^2 , chi square (degrees of freedom); CFI, comparative fit index; RMSEA root mean square error of approximation; CI, confidence interval; TLI, Tucker Lewis Index; AIC, Akaike's Information Criterion.

the chi-square difference test remained marginally significant (χ^2 difference (11) = 22.3, $p = 0.06$, CFI = 0.96, TLI = 0.95, RMSEA = 0.031 (0.28–0.35), AIC = 900.491). Thus, several items for anxiety and depression do not show invariance over time.

Cross-sex equality of only the constrained factor loadings in Model 3 was then examined (Model 4). The fit of this model was good (χ^2 (598) = 1023.2, CFI = 0.95, TLI = 0.93, RMSEA = 0.025 (0.22–0.27), AIC = 1516.427) although, again the chi-square difference test approached statistical significance ($p = 0.08$). Examination of factor loadings by sex from Model 4 (Table 3) does not suggest significant discrepancies in the pattern of loadings by sex, thus, we infer that the constrained items show invariance across sex.

Discussion

The purpose of the present study was to examine the stability of the factor structure of anxiety and depression symptoms across important developmental periods and across sex in a large nationally representative sample of children. Our primary hypothesis – that anxiety and depression would be represented as two stable factors invariant of time and sex – was supported with an important caveat. While the factors showed configural invariance, anxiety and depression items were not equivalent across time for boys and girls. This important finding suggests that while anxiety and depression can be measured independently of each other in childhood, individual items

for depression and anxiety may change in their strength as indicators of respective constructs across age.

We were unable to find previous references to anxiety or depression item variance across childhood in the current literature. The only other longitudinal studies which studied the factor structure of anxiety and depression across childhood (Lonigan *et al.*, 2003; Cole *et al.*, 1998) used previously validated scales (RCMAS and CDI) and did not examine the stability of the items within these scales over time. This finding makes intuitive sense since mothers are rating their children and we might expect that the subjective impression of the child as anxious, or depressed, by these characteristics is most salient in an older child.

Despite the possibility that some items may be more specific indicators of behaviors in children of different ages, studies whose aim is to measure changes in emotional or behavioral constructs in a developmentally sensitive way need to use items shown to be equally useful to represent problem behaviors across the ages studied. When this is not possible, as in our case, estimates of mean levels of anxiety and depression across varying ages are not likely accurately measured. In this study, given the pattern of factor loadings of these variable items increases over time, it is likely that average scores were overestimated in earlier ages, obscuring possible age differences.

Consistent with previous research in the topic area (Angold *et al.*, 1999), we found that anxiety and depression were highly correlated in each age group and in both boys and girls, suggesting that although they can be measured independently they have significant shared variance. Mean levels of anxiety and depression did not vary by sex whereas we would have expected levels to be significantly higher in girls than boys by Time 2 or Time 3. As the items were not invariant across age, this finding is difficult to interpret, however the most likely explanation for the absence of sex difference is that our oldest youth were 13 years and this difference had not yet emerged.

Although our findings require replication in future studies, it suggests that in the NLSY, such items are not equally representative of anxiety and depression across this age group and researchers using these items need to be aware of this.

Study limitations

We could not replicate a test of the “Tripartite” three-factor model in the present study because the items available were insufficient to approximate the physiologic arousal domain proposed by Clark and Watson (1991). Such item coverage could be incorporated into future epidemiologic studies as it has shown promise as a useful psychopathologic construct for differentiating various

Table 3. Maximum-likelihood estimates of the factor loadings of items for anxiety and depression in the final model

Item	Age 4–6+ (1994)		Age 8–10+ (1998)		Age 12–14+ (2002)	
	Boys	Girls	Boys	Girls	Boys	Girls
Anxiety						
1. Nervous	50	59	70	67	67	69
2. Fearful	51	58	72	69	63	59
3. Mind off	50	51	58	51	66	68
4. Dependent	50	46	55	50	55	51
5. Worried	54	55	71	62	67	64
Depression						
1. Not loved	52	50	57	54	63	62
2. Worthless	71	62	76	65	70	74
3. Sad	66	61	69	66	72	79
4. Withdrawn	53	33	56	43	46	56

All factor loadings are significant at $p < 0.001$.

anxiety phenotypes in children (Chorpita and Daleiden, 2002; McDewitt-Murphy *et al.*, 2007). However, given that the anxiety factor we studied is similar to a generalized anxiety construct we can infer that generalized anxiety and depression constructs can be differentiated in childhood and show measurement stability over time despite concurrent associations. This finding complements work in adults which suggests that Generalized Anxiety Disorder (GAD) and depression have variable orders of presentation in terms of longitudinal course and high rates of co-occurrence (Moffitt *et al.*, 2007) which are not the result of symptom overlap (Sunderland *et al.*, 2010). Other limitations were the availability of a small pool of items (BPI) and not scales previously validated against clinical diagnoses to predict the latent constructs representing internalizing disorders. This situation is often the case with data from large national datasets where items are chosen to reduce respondent burden. Previous research has shown that syndromes identified using behavior checklists may be analogous to those identified in clinical settings (Boyle *et al.*, 1997; Connor-Smith and Compas, 2003), however short scales have less utility for classification or differentiation of constructs (Emons *et al.*, 2007). We did select items that had acceptable face validity to clinical experts, were internally consistent ($\alpha = 0.65\text{--}0.77$) and empirically represented the chosen factor better than the alternative factor (anxiety versus depression, and vice-versa), to increase the validity of the factors. Test-retest reliability data is not available, and construct validity testing is sparse for the BPI scale.

Mothers were the sole informants in this study, limiting our ability to determine whether the patterned stability in measurement structure over time reflects stability in the constructs or mother's pattern of responding. The CFA permitted us to estimate the effect due to correlated errors across her responses, and indeed the magnitude of the estimated error variance was substantial, attesting to the importance of trying to parse it apart from an etiologic perspective, and the use of multiple raters or non-symptom methods would be required to answer this important question. Moreover, maternal responses were based on a three-point frequency scale which could have attenuated the correlations reported herein.

Given that the effective sample size changes significantly and degrees of freedom are increased when testing equivalence across time and sex simultaneously, it is

difficult to interpret whether there is also additional item variability by sex above that estimated to occur across time. Additional work is required to examine item stability across childhood, and such studies will need larger sample sizes to test sex and time invariance simultaneously.

Implications for future research

Given the limitations of reliability and validity of the BPI, it is prudent that researchers seeking to examine questions about children's psychological development using BPI data should ensure that they have a method to examine test-retest reliability, construct validity and measurement invariance before entering into longitudinal data analyses. As a general point, tests of measurement invariance of anxiety and depression factors are needed using additional epidemiologic samples and data types, as ours are the only ones to our knowledge.

There is evidence for a Tripartite Model, or some variant of it to describe relations between internalizing symptoms in childhood (Clark *et al.*, 1991; Mineka *et al.*, 1998; Anderson *et al.*, 2007). Future studies should include tests of a multi-dimensional anxiety construct, its measurement stability and relationships with depressive symptoms over time.

Demonstration of independent measurement structure does not permit researchers to infer validity of the anxiety and depression factors. Therefore whether children with high anxiety or depression scores have different risk factors or should be classified differently requires additional testing. As we have shown that these factors can be measured independently across childhood, models predicting anxiety and depression separately and in combination over time are likely to be valid, and those predicting "internalizing" disorders alone are likely insufficient.

Acknowledgments

The first author (KRB) wishes to acknowledge the financial support of a strategic training fellowship for doctoral studies provided by the Canadian Child Health Clinician Scientist Program in collaboration with the Canadian Institutes for Health Research.

Declaration of interest statement

The authors have no competing interests.

References

Achenbach T.M., Edelbrock C.S. (1981) Behavioral problems and competencies reported by parents of normal and disturbed children	aged four through sixteen. <i>Monograph of the Society for Research in Child Development</i>, 46, 1–82.	American Psychiatric Association (2001) <i>Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition - Text Revisions</i>

- (*DSM IV-TR*), Washington, DC, American Psychiatric Press.
- Anderson E.R., Hope D.A. (2007) A review of the tripartite model for understanding the link between anxiety and depression in youth. *Clinical Psychology Review*, **25**, 278–287.
- Angold A., Costello E.J., Erkanli C. (1999) Psychiatric comorbidity in adolescents: empirical, theoretical, and methodological issues. *The American Journal of Psychiatry*, **150**, 1779–1791.
- Arbuckle J.L. (1996) Full information estimation in the presence of incomplete data. In Marcoulides G.A., Schumacker R.E. (eds) *Advanced Structural Equation Modeling: Issues and Techniques*, pp. 243–277, NJ, Erlbaum: Mahwah.
- Arbuckle J.L. (2006) Amos (Version 7.0) [Computer Program], Chicago, SPSS.
- Axelson D.A., Birmaher, B. (2001) Relation between anxiety and depressive disorders in childhood and adolescence. *Depression and Anxiety*, **14**, 67–78.
- Bentler P.M. (2000) Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238–246.
- Boyle M.H., Offord D.R., Racine Y.A., Szatmari P., Sanford M., Fleming J.E. (1997) Adequacy of interviews vs. checklists for classifying childhood psychiatric disorder based on parent reports. *Archives of General Psychiatry*, **54**, 793–799.
- Byrne B.M. (2004) Testing for multigroup invariance using AMOS Graphics: a road less travelled. *Structural Equation Modeling*, **11**, 272–300.
- Center for Human Resource Research (2002) *NLSY79 Child and Young Adult Data Users Guide: A Guide to the 1986–2000 Child Data, 1994–2000 Young Adult Data*, Columbus, OH, Ohio State University.
- Chorpita B.F., Daleiden E.L. (2002) Tripartite dimensions of emotion in a child clinical sample: measurement strategies and implications for clinical utility. *Journal of Consulting and Clinical Psychology*, **70**, 1150–1160.
- Chorpita B.F., Daleiden E.L., Moffitt C., Yim L., Umemoto L.A. (2000) Assessment of tripartite factors of emotion in children and adolescents. I: Structural validity and normative data of an affect and arousal scale. *Journal of Psychopathology and Behavioral Assessment*, **22**, 141–160.
- Cicchetti D., Toth S.L. (1998) The development of depression in children and adolescents. *American Psychologist*, **53**, 221–241.
- Clark L.A., Watson D. (1991) Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, **100**, 316–336.
- Colder C.R., Mott J.A., Berman A.S. (2002) Interactive effects of infant activity level and fear on growth trajectories of early childhood behaviour problems. *Developmental Psychopathology*, **14**, 1–23.
- Cole D.A., Truglio R., Peeke L. (1997) Relation between symptoms of anxiety and depression in children: a multitrait–multimethod–multigroup assessment. *Journal of Consulting and Clinical Psychology*, **65**, 110–119.
- Cole D.A., Peeke L.G., Martin J.M., Truglio R., Seroczynski A.D. (1998) A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology*, **66**, 451–460.
- Connor-Smith J.K., Compas B.E. (2003) Analog measures of DSM-IV mood and anxiety disorders based on behaviour checklists. *Journal of Psychopathology and Behavioral Assessment*, **27**, 35–48.
- Costello E.J., Egger H.L., Angold A. (2004) Developmental epidemiology of anxiety disorder. In Ollendick T.H., March J.S. (eds) *Phobic and Anxiety Disorders in Children and Adolescents: A Clinician's Guide to Effective Psychosocial and Pharmacological Interventions*, pp. 61–91, New York, Oxford University Press.
- Duffy C.J., Cunningham E.G., Moore S.M. (2005) Brief report: the factor structure of mood states in an early adolescent sample. *Journal of Adolescence*, **28**, 677–680.
- Emons W.H., Sijsma K., Meijer R.R. (2007) On the consistency of individual classification using short scales. *Psychological Methods*, **12**, 105–120.
- Hankin B.L., Abramson L.Y., Moffitt T.E., Silva P.A., McGee R., Angell K.A. (1998) Development of depression from preadolescence to young adulthood: emerging gender differences in a 10 year longitudinal study. *Journal of Abnormal Psychology*, **107**, 128–141.
- Huebner E., Dew T. (1995) Preliminary validation of the Positive and Negative Affect Schedule in adolescents. *Journal of Psychoeducational Assessment*, **13**, 286–293.
- Jaccard J., Wan C.K. (1995) Measurement error in the analysis of interaction effects between continuous predictors using multiple regression – multiple indicator and structural equation approaches. *Psychological Bulletin*, **117**, 348–357.
- Joiner T.E. Jr, Catanzaro S.J., Laurent J. (1996) Tripartite structure of positive and negative affect, depression, and anxiety in child and adolescent psychiatric inpatients. *Journal of Abnormal Psychology*, **105**, 401–409.
- Kovacs M. (1985) The Children's Depression Inventory. *Psychopharmacology Bulletin*, **21**, 995–998.
- Kovacs M., Gatsonis C., Paulauskas S.L., Richards C. (1989) Depressive disorders in childhood. IV. A longitudinal study of comorbidity with and risk for anxiety disorders. *Archives of General Psychiatry*, **46**, 776–782.
- Laurent J., Ettelson R. (2001) An examination of the tripartite model of anxiety and depression and its application to youth. *Clinical Child and Family Psychology Review*, **4**, 209–230.
- Laurent J., Catanzaro S.J., Joiner T.E. Jr (2004) Development and preliminary validation of the physiological hyperarousal scale for children. *Psychological Assessment*, **16**, 373–380.
- Lee L.-C., Reebok G.W. (2002) Anxiety and depression in children: A test of the positive and negative affect model. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 429–416.
- Lewinsohn P.M., Gotlib I.H., Lewinsohn M. (1998) Gender differences in anxiety disorders and anxiety symptoms in adolescents. *Journal of Abnormal Psychology*, **107**, 109–117.
- Little T.D., Rubin D.B. (1989) The analysis of social science data with missing values. *Sociological Methods & Research*, **18**, 292–326.
- Lonigan C.J., Phillips B.M., Hooe E.S. (2003) Relations of positive and negative affectivity to anxiety and depression in children: evidence from a latent variable longitudinal study. *Journal of Consulting and Clinical Psychology*, **71**, 465–481.
- McDevitt-Murphy M.E., Weathers F.W., Flood A.M., Eakin D.E., Benson T.A. (2007) The utility of the PAI and the MMPI-2 for discriminating PTSD, depression, and social phobia in trauma-exposed college students. *Assessment*, **14**, 181–195.
- Miles J., Shevlin M. (2007) A time and a place for incremental fit indices. *Personality and Individual Differences*, **42**, 869–874.
- Mineka S., Watson D., Clark L.A. (1998) Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, **49**, 377–412.
- Moffitt T.E., Harrington H., Caspi A., Kim-Cohen J., Goldberg D., Gregory A.M., Poulton, R. (2007) Depression and generalized anxiety disorder: cumulative and sequential comorbidity in a birth

- cohort followed prospectively to age 32 years. *Archives of General Psychiatry*, **64**, 651–660.
- Nunally J.C., Bernstein I.H. (1994) *Psychometric Theory*, New York, McGraw-Hill.
- Peterson J.L., Zill N. (1986) Marital disruption, parent–child relationships, and behavior problems in children. *Journal of Marriage and Family*, **48**, 295–307.
- Radloff L. (1997) The CES-D Scale: a self report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 385–401.
- Reynolds, C.R. (1980) Concurrent validity of what i think and feel: the Revised Children’s Manifest Anxiety Scale. *Journal of Consulting and Clinical Psychology*, **48**, 774–775.
- Shafer J.L., Graham J.W. (2002) Missing data: our view of the state of the art. *Psychological Methods*, **6**, 147–177.
- Sunderland M., Mewton L., Slade T., Baillie A.J. (2010) Investigating differential symptom profiles in major depressive episode with and without generalized anxiety disorder: true co-morbidity or symptom similarity? *Psychological Medicine*, **40**, 1113–1123.
- Watson D., Clark L.A., Tellegen A. (1988) Development and validation of a brief measure of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, **54**, 1063–1070.
- Zill N., Snyder N. (1981) *Tabulations from Public Use Data of 1981 Child Health Supplement to the National Health Interview Survey*, Washington, DC, Child Trends.