Measurement Invariance Across Parent and Self-Ratings of Extremely Low Birth Weight Survivors and Normal Birth Weight Controls in Childhood and Adolescence on the Child Behavior Checklist and Youth Self-Report

Ryan J. Van Lieshout,1 MD, PHD, Michael H. Boyle,1 PHD, Louis A. Schmidt,2 PHD, Saroj Saigal,3 MD, and Mark A. Ferro,1,3 PHD

1Department of Psychiatry and Behavioural Neurosciences, 2Department of Psychology, Neuroscience and Behaviour, and 3Department of Pediatrics, McMaster University

All correspondence concerning this article should be addressed to Mark Ferro, PHD, Departments of Psychiatry & Behavioural Neurosciences and Pediatrics, McMaster University, Offord Centre for Child Studies, Chedoke Site, Central Building, Room 304, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada. E-mail: ferroma@mcmaster.ca

Received December 24, 2014; revisions received April 13, 2015; accepted April 20, 2015

Abstract

Objective This study examined the measurement invariance of the Child Behavior Checklist (CBCL) and Youth Self-Report (YSR) DSM-oriented scales between extremely low birth weight (ELBW) and normal birth weight (NBW) youth. Methods The sample included 158 ELBW survivors and 145 matched, NBW controls at 8 and 12–16 years of age. Results Strict invariance was established at 8 years for parent-reported CBCL attention-deficit hyperactivity, conduct, and oppositional defiant scales, though invariance could not be established for affective problems at 8 or 12–16 years. Strict invariance was observed between 12–16-year-old ELBW and NBW groups on attention-deficit hyperactivity, anxiety, and oppositional defiant CBCL and YSR scales. Invariance could not be established for youth-reported conduct problems. Conclusions While the majority of CBCL/YSR DSM-oriented subscales assess the same concepts in both ELBW and NBW children and adolescents across parent and youth reports, this may not be the case for affective and conduct problems.

Key words: adolescent; child; Child Behavior Checklist; extremely low birth weight; measurement invariance; youth self-report.
Because children born prematurely are exposed to a wide range of adverse intrauterine and early postnatal environmental exposures, it is not surprising that they have neuroanatomical abnormalities not seen in term-born children (Volpe, 2009). These include neuronal and axonal pathology affecting specific white-matter regions (Volpe, 2008), as well as the thalamus (Pierson et al., 2007), basal ganglia (Inder, Warfield, Wang, Huppi, & Volpe, 2005), and cerebral cortex (Peterson et al., 2003). Alterations in all of these regions have been implicated in the pathogenesis of psychiatric problems, but in combination they may lead to unique phenotypes that may be difficult to detect using measures of psychopathology developed and normed using general population samples. For example, in childhood, ELBW and VLBW survivors have been observed to develop symptoms of an inattentive form of ADHD (Botting, Powls, Cooke, & Marlow, 1997; Elgen, Sommerfeldt, & Markestad, 2002; Indredavik et al., 2004; Szatmari et al., 1993).

Another factor that can affect ratings of psychopathology in preterm and term-born children is the identity of the informant who provides them. Discrepancies frequently exist between different informants’ (e.g., youth, parents, teachers) assessments of child and adolescent psychopathology (De Los Reyes & Kazdin, 2005). These can be related to the disorder of interest, characteristics of the child (e.g., age, gender), and the identity of the informant. Of particular relevance to ELBW and VLBW children are the characteristics of the parent(s) who rate their child’s behavior. Having an infant who is born preterm can be a difficult experience for families that are associated with high levels of stress (Singer et al., 2007). Moreover, factors such as the infant’s postnatal complications and extended hospital stays can have a lasting impact on parents in the form of depression, anxiety, and altered parent–child interactions (Singer, Salvador, Guo, Collin, & Baley, 1999). Such factors may differentially influence ratings of psychopathology in ELBW youth and lead to erroneous conclusions about risk.

Studies that have compared rates of emotional and behavioral problems in ELBW survivors and normal birth weight (NBW) controls raise questions about the cross-group validity of the scales used to define psychopathology. The importance of establishing the validity of these measures across those born premature and at term is highlighted by the fact that the two groups may differ in the way psychiatric problems develop or manifest and how different informants perceive these individuals and their difficulties. Traditional means of assessing reliability and validity cannot establish if informants reporting on psychopathology in ELBW and NBW youth interpret scales or scale items in conceptually similar ways. If a behavior scale or the items that comprise it have a different meaning or manifest differently in one group compared with another, the application of standard cutoffs can lead to misclassification, invalid group comparisons, and incorrect conclusions about risk, potentially undermining an entire scientific literature.

Measurement invariance concerns the extent to which the psychometric properties of observed scale items are generalizable across groups (Brown, 2006; Vandenberg & Lance, 2000). Its presence signifies that the same construct is being measured in different groups. Demonstrating measurement invariance makes it possible to interpret differences between groups as real and meaningful. Without evaluating measurement invariance, researchers cannot be certain if observed differences reflect true differences or whether they represent differential interpretation of items or structure of the construct (Brown, 2006). Despite evidence of response heterogeneity, most research takes measurement invariance for granted (Muthén, 1989).

The most commonly used measure of psychopathology in children and adolescents born at ELBW is the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983). Indeed, of the 24 studies that comprised the aforementioned systematic reviews, 14 defined emotional and behavioral problems using the CBCL (Aarnoudse-Moens et al., 2009; Bhutta et al., 2002). As a result, the fundamental basis of our understanding of the psychiatric risk associated with being born prematurely is based on this measure.

The parent-reported CBCL and the youth-reported Youth Self-Report (YSR; Achenbach, 1991) measure are dimensional instruments scored on eight statistically derived syndrome scales (often referred to as empirical scales) and six DSM-oriented scales (affective, anxiety, somatic, ADHD, oppositional defiant, and conduct problems; Achenbach, Dumenci, & Rescorla, 2003). These measures are widely used and have available norms throughout childhood and adolescence (Achenbach, 1991; Achenbach & Edelbrock, 1983; Achenbach & Rescorla, 2001), and across genders and a range of cultures (Ivanova et al., 2007). While confirmatory factor analysis has established configural invariance in children across cultures (Ivanova et al., 2007), only recently have more stringent tests of invariance been applied to the CBCL. However, to our knowledge, only a few studies have examined the invariance of the CBCL beyond the configural level. Fonseca-Pedrero, Sierra-Baigrie, Lemos-Giraldez, Paino, & Muniz (2012) demonstrated that the YSR is invariant across gender and age in a large nonclinical sample of adolescents, while Gross et al. (2006) showed that the CBCL internalizing and externalizing scales are invariant across race/ethnicity (Black, Latino, non-Latino White), income
level, and language (Spanish vs. English). Recent work by Yarnell et al. (2013) reported that the internalizing and externalizing scales of the CBCL are invariant across different ethnic groups within a non-Western sample of 11-year-old children. Indeed, this was the first study to assess and demonstrate strict invariance of any of the CBCL scales in any population. Ferro, Boyle, Scott, & Dingle (2014) have also shown that the CBCL attention and YSR thought problems scales are invariant between adolescents with a chronic neurological problem (epilepsy), suggesting that the CBCL and YSR may be capable of reliably and validly assessing emotional and behavioral problems in adolescents with brain insults.

Despite the fact that the majority of the work examining the mental health of ELBW survivors and NBW controls has used the CBCL, and our understanding that ELBW survivors are at higher risk for certain emotional and behavioral problems is largely based on this measure, no study has tested for measurement invariance between ELBW and NBW groups. Establishing measurement invariance of the CBCL and YSR in ELBW and NBW groups will not only permit valid comparisons of scale scores between groups, but also allows for the defensible examination of the true prevalence of emotional and behavioral problems in ELBW survivors, as well as meaningful exploration of the risk factors for these difficulties (Gregorich, 2006). Finally, it will increase our confidence in previous reports that have suggested that ELBW survivors are at elevated risk of certain types of psychopathology and support the appropriateness of its use in future research.

Given this background, the objectives of this study were to: (1) assess the measurement invariance of parent reports on the CBCL DSM-oriented scales across ELBW survivors and NBW controls at age 8 and 12–16 years, and (2) to determine if measurement invariance could be established between these birth weight groups using self-reports on the DSM-oriented scales of the YSR at age 12–16 years. We hypothesized that differences in neuropathology, phenomenology, and parental reporting would result in a lack of measurement invariance on these scales in childhood and adolescence between ELBW survivors and NBW controls.

**Methods**

**Study Sample**

Participants in this study are members of the McMaster Extremely Low Birth Weight Cohort, a group of 179 ELBW survivors and 145 NBW controls born between 1977 and 1982 at the McMaster University Regional Perinatal Centre located in Hamilton, Ontario, Canada, to residents of central-west Ontario, a region consisting of Hamilton-Wentworth, Brant, Haldimand-Norfolk, Halton, Niagara, Waterloo, and Wellington counties. The original purpose of the study was to determine the immediate and long-term developmental and health outcomes of infants born 501–1,000 g in the McMaster Health Region. ELBW participants (birth weight: 501–1,000 g) comprised 397 infants recruited at birth and have been followed longitudinally. Of the ELBW survivors, 179/397 (45%) survived to hospital discharge and 10 children subsequently died, leaving 169 survivors at age 8 years. NBW controls were >2,500 g at birth, recruited when the ELBW group was 8 years old, and comprised 145 children born between 1977 and 1982. Control subjects were selected from a random sample of students in the Hamilton Public and Catholic School Systems who were born at term and were matched with the ELBW survivors on age, sex, and social class (Saigal, Szatmari, Rosenbaum, & Campbell, 1991). No inclusion or exclusion criteria were applied outside of the aforementioned birth weight cutoffs. Of these, 158 ELBW survivors and 145 NBW participants contributed data for this study. The McMaster University Health Sciences Research Ethics Board approved all study procedures.

**Measures**

The CBCL and YSR are measures commonly used to assess behavioral problems in children and adolescents and their validity and reliability have been extensively documented (Achenbach & Rescorla, 2001). To address the objectives of this study, the DSM-oriented scales of ADHD, affective problems, anxiety problems, conduct problems, and oppositional defiant disorder derived from the CBCL and YSR were tested for measurement invariance (Achenbach et al., 2003). The DSM-oriented scales were selected over the CBCL and YSR empirical scales because the questionnaires that were originally used in this cohort study contained the complete set of items comprising the DSM scales, but far fewer of the items needed to create the CBCL and YSR empirical scales. As a result, our data did not permit us to adequately capture the constructs reflected in the empirically derived scales.

Parents reported on their child’s behavior when children were 8 and 12–16 years of age. YSR data were obtained at 12–16 years of age. The 8-year-old follow-up occurred between 1985 and 1990 and adolescent data were collected 4–8 years later. Data on the CBCL and YSR were coded such that higher scores were reflective of more symptoms of problem behavior with the following response options: 0 = “Not true”; 1 = “Somewhat or sometimes true”; and 2 = “Very true or often true.”
**Procedure**

Confirmatory factor analysis was used to examine the measurement invariance of the CBCL and YSR DSM scales in which a hierarchical set of equality constraints between ELBW and NBW controls was specified. This sequential testing and model specification strategy was devised from published guidelines for establishing measurement invariance of factor models based on ordered categorical items (Millsap & Yun-Tein, 2004): (i) configural invariance imposes no equality constraints on parameters and was used as the origin for more complex models to be tested (Brown, 2006); (ii) strong invariance (i.e., constrained factor loadings and item thresholds) examines the extent to which the magnitude of the factor loadings (A) for particular items are the same between groups and is a prerequisite for making valid comparisons (Bollen, 1989), whereas invariance of item thresholds (v) verifies whether mean differences at the item level are fully explained by mean differences at the factor level; and (iii) strict invariance (i.e., constrained residual variances) is performed to determine whether the variances (θ) of the regression equations for each item are equivalent between groups and are required for defensible item-score comparisons (i.e., average item scores) between groups (Steinmetz, Schmidt, Tina-Booh, Wieczorek, & Schwartz, 2009). Owing to the ordered categorical nature of responses for the CBCL and YSR, the confirmatory factor model was estimated with a weighted least squares estimator using a diagonal weight matrix (Muthén & Muthén, 2010). This estimator uses pairwise deletion to account for missing data, which efficiently generates robust unbiased parameter estimates (Asparouhov & Muthén, 2010).

**Statistical Analysis**

Measurement invariance was considered present when, after imposing a constraint, there was no appreciable worsening of model fit. If this condition was met, invariance testing proceeded to the application of the next constraint. If there was significant worsening of fit, modification indices were reviewed. If the output suggested that removing constraints on relevant noninvariant parameters would improve model fit, these constraints were removed. This re-specified model was then tested against the less constrained model and is known as *partial* invariance (Byrne, Shavelson, & Muthén, 1989). Researchers have argued that *full* measurement invariance is not required for substantive analyses if at least a subset of parameters is determined to be invariant. Recent work by Steinmetz (2013) has shown that unequal factor loadings have only minimal effects on interpretations of latent mean differences, whereas unequal intercepts/thresholds substantially influence inferences on latent mean differences. As a result, modifications to improve model fit were limited to factor loadings only. If adequate model fit was not achieved at the configurational level, tests of measurement invariance were not conducted.

Determination of model fit was based on two goodness-of-fit indices. These were the Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) and 90% confidence interval (Cheung & Rensvold, 2002). Adequate model fit was defined using the following cutoffs: CFI > 0.95 or RMSEA < 0.06 (Cheung & Rensvold, 2002). Invariance was established if the change in model fit met the following thresholds: ΔCFI ≤ −0.010 or ARMSEA ≤ 0.015 (Chen, 2007). The χ^2 goodness-of-fit and Δχ^2 were included for completeness, but decisions of model fit and invariance were not based on these indices. Instead, model fit or change in model fit must have satisfied at least one of the CFI or RMSEA criteria. Descriptive statistics and comparisons between the ELBW and NBW samples were calculated using SAS 9.2 (SAS Institute Inc., United States). Analyses associated with measurement invariance testing were performed with Mplus 6.11 (Muthén & Muthén, United States).

**Missing Data**

A total of 21 participants (6.5%) had completely missing data for the CBCL and YSR and were removed from the analysis. Of the remaining 303 participants, 96 (31.7%) had one or two missing response items on the CBCL or YSR at the 8-year and 12–16-year follow-up. Fifty-nine (19.5%) had ≥3 missing items. Missing data were not associated with birth weight (odds ratio [OR] = 1.39 [0.88, 2.18]; p = .156), gestational age (OR = 0.83 [0.65, 1.06]; p = .144), or socioeconomic status (OR = 1.10 [0.87, 1.39]; p = .422).

**Results**

**Sample Characteristics**

One hundred fifty-eight ELBW survivors and 145 NBW participants contributed data for this study. The mean age at follow-up at the two measurement occasions was 8 and 14.2 years. Mean birth weight was 836.0 g (124.5) and 3,372.8 g (488.6) for ELBW and NBW groups, respectively. This difference was statistically significant (p < .001). Not surprisingly, ELBW and NBW groups also differed on the mean gestational age at which they were born (p < .001). These

---

1 To achieve model identification, certain parameters must be fixed: The factor mean is fixed to zero and residual variances are fixed to one in the referent group (NBW); the factor loading of one item (i.e., referent item) is fixed to one in both groups; and, the first two thresholds for the referent item and the first threshold from the remaining items are fixed to equality between groups.
groups did not differ on mean age at testing, socioeconomic status, parental years of education, or the proportion raised in a lone parent family. Immigrant family status (those where both parents were born outside of Canada) did not differ significantly between ELBW and NBW groups (Table I). Thirteen percent of ELBW participants were small for gestational age and 15% were exposed to a complete course of prenatal steroids.

CBCL Measurement Invariance at 8 Years

Attention-Deficit Hyperactivity (Six Items)

Model fit for ADHD was initially adequate ($\chi^2(24) = 58.85$, $p < .001$; CFI = 0.961; RMSEA = 0.099 [0.067, 0.131]) (Parent-report, Supplementary Table S1). Strong invariance was established by removing the constraints on the factor loading for Q3-fidgets: ($\Delta$CFI = 0.001; $\Delta$RMSEA = −0.009). Removal of this constraint was informed by previous findings of neuroanatomical abnormalities observed in ELBW children (Volpe, 2009), as well as evidence of inattentiveness in this group (Indredavik et al., 2004). Subsequently, strict invariance was also established ($\Delta$CFI = −0.001; $\Delta$RMSEA = −0.006).

Affective Problems (11 Items)

Items Q3-harms self and Q10-talks suicide were removed because of empty cells in their bivariate table resulting in model nonconvergence. An adequate model for affective problems was not obtained at the configural level ($\chi^2(59) = 184.33$, $p < .001$; CFI = 0.857; RMSEA = 0.080 [0.062, 0.097]) and testing did not proceed further.

Anxiety Problems (Six Items)

Model fit for anxiety problems was adequate at the configural level ($\chi^2(24) = 37.70$, $p = .037$; CFI = 0.957; RMSEA = 0.062 [0.015, 0.098]). Strong invariance was established ($\Delta$CFI = 0.006; ARMSEA = −0.008), but not strict invariance ($\Delta$CFI = −0.025; $\Delta$RMSEA = 0.010).

Conduct Problems (10 Items)

Items relating to truancy/skipping school, vandalism, running away from home, and setting fires were removed before specifying the model because of no variation in responses for both the ELBW and NBW groups. Items Q1-mean to animals and Q10-steals outside home were removed owing to collinearity resulting in a nonidentified model. Subsequent model fit was good ($\chi^2(77) = 113.24$, $p = .005$; CFI = 0.963; RMSEA = 0.056 [0.032, 0.078]) and strict invariance was obtained ($\Delta$CFI = 0.010; $\Delta$RMSEA = −0.009).

Oppositional Defiant Disorder (Five Items)

Model fit for oppositional defiant disorder was initially adequate ($\chi^2(15) = 56.02$, $p < .001$; CFI = 0.950; RMSEA = 0.136 [0.099, 0.175]). Strong invariance was established by removing the constraints on the factor loading for Q3-disobeys at school ($\Delta$CFI = 0.001; $\Delta$RMSEA = −0.014). Justification for removing this constraint was informed by previous findings of inattentiveness among ELBW children (Indredavik et al., 2004) that may be captured by disobedience at school. Subsequently, strict invariance was also established ($\Delta$CFI = 0.002; $\Delta$RMSEA = −0.014).

CBCL Measurement Invariance at 12–16 Years

Attention-Deficit Hyperactivity (Nine Items)

Adequate model fit for ADHD was adequate at the configural level ($\chi^2(63) = 123.21$, $p < .001$; CFI = 0.975; RMSEA = 0.085 [0.062, 0.107]). Strict invariance was established ($\Delta$CFI = −0.001; $\Delta$RMSEA = −0.002) (Parent-report, Supplementary Table S2).

Affective Problems (11 Items)

An adequate model for affective problems was not obtained at the configural level ($\chi^2(97) = 181.71$, Table I. Characteristics of Participants and Their Families of Origin at Age 8

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ELBW (n = 158)</th>
<th>NBW (n = 145)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male, %</td>
<td>45.6</td>
<td>45.5</td>
<td>.99</td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
<td>7.9 (0.43)</td>
<td>8.0 (0.44)</td>
<td>.18</td>
</tr>
<tr>
<td>Mean birth weight (SD)</td>
<td>836 (124)</td>
<td>3373 (489)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mean gestational age (SD)</td>
<td>27 (2.3)</td>
<td>40 (0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lone-parent family, %</td>
<td>12.0</td>
<td>12.4</td>
<td>.99</td>
</tr>
<tr>
<td>Immigrant family, %</td>
<td>13.3</td>
<td>17.2</td>
<td>.87</td>
</tr>
<tr>
<td>Family socioeconomic class (quintiles), %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>7.6</td>
<td>3.9</td>
<td>.52</td>
</tr>
<tr>
<td>Second</td>
<td>18.6</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>Third</td>
<td>37.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>Fourth</td>
<td>29.7</td>
<td>36.6</td>
<td></td>
</tr>
<tr>
<td>Fifth</td>
<td>6.9</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Mean maternal education, years (SD)</td>
<td>13.8 (2.62)</td>
<td>14.3 (2.63)</td>
<td>.09</td>
</tr>
<tr>
<td>Mean paternal education, years (SD)</td>
<td>14.1 (3.01)</td>
<td>14.1 (2.67)</td>
<td>.88</td>
</tr>
</tbody>
</table>
Anxiety Problems (Six Items)
Model fit for anxiety problems was excellent ($\chi^2(23) = 24.98, p = .352; \text{CFI} = 0.995; \text{RMSEA} = 0.025 [0.000, 0.077]$). Strict invariance was established ($\Delta\text{CFI} = 0.000; \Delta\text{RMSEA} = 0.000$).

Conduct Problems (Nine Items)
Items relating to truancy/skipping school and vandalism were removed owing to empty cells in their bivariate table resulting in model nonconvergence. Items Q2-steals inside home, Q6-lies/cheats, Q10-hangs out with troublesome kids, and Q13-fights were removed owing to collinearity resulting in a nonidentified model. Subsequent model fit was good ($\chi^2(62) = 155.29, p < .001; \text{CFI} = 0.954; \text{RMSEA} = 0.105 [0.085, 0.126]$), but strong invariance was not obtained ($\Delta\text{CFI} = -0.010; \Delta\text{RMSEA} = 0.004$).

Oppositional Defiant Disorder (Five Items)
Model fit for oppositional defiant disorder was adequate ($\chi^2(15) = 28.57, p = .018; \text{CFI} = 0.989; \text{RMSEA} = 0.082 [0.033, 0.128]$) and strict invariance was achieved ($\Delta\text{CFI} = -0.006; \Delta\text{RMSEA} = 0.013$).

YSR Measurement Invariance at 12–16 Years
Attention-Deficit Hyperactivity (12 Items)
Data fit the model well for ADHD ($\chi^2(120) = 162.76, p = .006; \text{CFI} = 0.973; \text{RMSEA} = 0.053 [0.029, 0.072]$) and subsequently, strict invariance was established ($\Delta\text{CFI} = -0.008; \Delta\text{RMSEA} = 0.005$) (Youth-Report, Supplementary Table S3).

Affective Problems (12 Items)
Configural model fit was adequate ($\chi^2(120) = 185.12, p < .001; \text{CFI} = 0.968; \text{RMSEA} = 0.066 [0.046, 0.083]$). Strict invariance was demonstrated ($\Delta\text{CFI} = -0.009; \Delta\text{RMSEA} = 0.006$).

Anxiety Problems (Six Items)
Data fit the model very well for anxiety problems ($\chi^2(24) = 22.72, p = .537; \text{CFI} = 1.000; \text{RMSEA} = 0.000 [0.000, 0.067]$) and strict invariance was established ($\Delta\text{CFI} = 0.000; \Delta\text{RMSEA} = 0.000$).

Conduct Problems (14 Items)
Items relating to being mean to others and animals were removed before specifying the model because of no variation in responses for both the ELBW and NBW groups. An adequate model for conduct problems was not obtained at the configural level ($\chi^2(164) = 249.06, p < .001; \text{CFI} = 0.944; \text{RMSEA} = 0.063 [0.047, 0.079]$) and testing did not proceed further.

Oppositional Defiant Disorder (Five Items)
Model fit for oppositional defiant disorder was adequate ($\chi^2(15) = 26.05, p = .038; \text{CFI} = 0.984; \text{RMSEA} = 0.076 [0.018, 0.123]$). Strict invariance was achieved ($\Delta\text{CFI} = -0.001; \Delta\text{RMSEA} = -0.004$).

Discussion
In this study, we showed that strict measurement invariance is present across 8-year-old ELBW survivors and NBW controls on the parent-reported CBCL DSM-oriented ADHD and oppositional defiant disorder scales. Strict invariance was also found for the conduct problems scale, but this occurred only after removal of six of its original 16 items. Strong invariance was found for anxiety problems, but invariance could not be demonstrated for the affective problems scale. We were also able to establish strict invariance for parental ratings of 12–16-year-olds on all CBCL DSM-oriented scales with the exception of the affective and conduct problems scales. While strong invariance was demonstrated for the conduct problems scale (after removal of six items), the affective scale was noninvariant between ELBW survivors and NBW controls. Finally, strict invariance was established for all self-reported YSR scales at 12–16 years of age except for the conduct problems scale. As a result, our hypotheses were partially supported, particularly as they related to the affective and conduct problems scales.

In this study, invariance was observed across factor loadings and item thresholds (strong invariance) in all but the affective problems scale according to parental reports and the conduct scale as reported by youth. These findings suggest that the degree of bias of the CBCL and YSR was equal between groups for most scales. Strict invariance suggested that the items had the same accuracy and were functionally reliable for both ELBW and NBW participants. With the exception of the aforementioned scales, our results suggest that the CBCL and YSR generally function similarly between ELBW survivors and NBW controls throughout childhood and adolescence and across parent reports and YSRs. However, further study is required to replicate our results, particularly those pertaining to the affective and conduct problems scales.

These findings are important because the CBCL provides much of the basis of our understanding of the psychiatric risks in ELBW survivors during childhood and adolescence. Substantively, since the most consistently reported differences between ELBW survivors and NBW controls in children and adolescents are those in the areas of ADHD and anxiety, our results suggest that the findings of many of the 14 studies that have used the CBCL in studies of premature infants and term-born controls are based on a measure that assesses the same concepts in both of these populations and are valid.
This study is one of the few that has examined the measurement invariance of the CBCL in general, and, to our knowledge, is the first to assess it throughout childhood and adolescence, demonstrate that this is consistent across informants in adolescence, and to examine measurement invariance in an at-risk population exposed to significant early adversity. This work also adds to the accumulating literature on the measurement invariance of the CBCL across different populations. In particular, our work is consistent with previous findings of configural invariance across age and gender (Fonseca-Pedrero et al., 2012), race/ethnicity (Gross et al., 2006; Yarnell et al., 2013), and physical health status (Ferro et al., 2014) with the CBCL and YSR, and suggests that these measures may be capable of reliably and validly assessing emotional and behavioral problems in adolescents with a range of neuropathology.

Measurement invariance was not present in parental reports of affective problems in ELBW survivors and NBW controls at either 8 or 12–16 years of age. It is not clear why parents of ELBW and NBW youth systematically differ on reports of their children’s affective problems. Delivering and raising a child born preterm can be an extremely stressful experience requiring significant investments of time, energy, and financial resources and may affect parents’ moods, as well as their perceptions of their own children’s feelings. Perhaps parents of ELBW survivors may be less aware of the depressive symptoms of their offspring because they are less prominent than the health and social problems ELBW children tend to manifest. The ongoing health challenges facing these children may particularly impact parents’ reports on the seven physical symptoms that contribute to this 14-item scale. The implications of this finding for previous research in this field are unclear. While the CBCL DSM-oriented affective problems scale has not been widely used, it shares a number of items with the empirical anxious/depressed scale. And while the ELBW “phenotype” has tended to be more withdrawn, the anxious/depressed scale is one of three that comprise the broadband internalizing scale. While our finding does not invalidate research based on this internalizing scale in youth, future work using the CBCL in this population may benefit from the establishment of measurement invariance between birth weight groups when parent reports are used.

Despite the fact that we were able to establish strong and strict invariance for the parent-reported CBCL conduct problems scale at age 8 and 12–16 years, respectively, this was only possible after a significant number of scale items were removed from each. As a result, the scales composed of the remaining items may not accurately reflect the originally intended construct. Moreover, measurement invariance could not be established for the self-reported YSR conduct problems scale at 12–16 years. Why ELBW and NBW youth systematically differ on their own reports of conduct problems is not known. Substantively, ELBW survivors tend to have a more shy, cautious, and reticent personality style that is somewhat incongruent with externalizing psychopathology (Hack et al., 2002, 2004; Schmidt, Miskovic, Boyle, & Saigal, 2008; Waxman, Van Lieshout, Saigal, Boyle, & Schmidt, 2013). Perhaps earlier manifestations of ELBW survivors’ views of themselves as being more shy, conscientious, and agreeable in late adolescence (Allin et al., 2006; Pesonen et al., 2008), a tendency to respond in a more socially desirable way (Allin et al., 2006) or higher rates of parental monitoring affect how they report on their own behavior in the realm of conduct problems. Given these findings, it is possible that noninvariance has contributed to the mixed results of studies reporting conduct-related behavior via YSRs as this work has shown elevated (Georgsdottir et al., 2013), reduced (Hallin & Stejnqvist, 2011), and equivalent levels (Saigal et al., 2003) of these problems in ELBW and NBW control subjects. Regardless, as strong invariance is sufficient for substantive comparisons using these scales, results based on parent reports likely remain valid. However, in future studies, invariance should be established before the acceptance of results based on youth reports of conduct problems in studies of preterm and term-born individuals.

Despite the generally positive nature of our findings for the field, it is important that our work be interpreted in the context of the following limitations. We used the CBCL and YSR DSM-oriented scales rather than the more commonly used empirical scales owing to item inclusion in our study at the child and adolescent waves. Unfortunately, up to 50% of the empirical scale items were missing for the majority of these scales and so a decision was made to examine the more complete DSM scales. While the DSM scales are likely clinically relevant, even though there is significant overlap between the DSM and empirical scales, it will be important that our findings are confirmed using the CBCL and YSR empirical scales. Likewise, in some instances, particularly the conduct disorder scales, several items were removed to achieve model identification/convergence. As a result, the original scale may not have been represented in our analyses. Moreover, because our sample contained only ELBW (and not VLBW) survivors, replication of our results is required in groups of individuals born less prematurely. It is also important to point out that our sample was born at a time when the survival rate of ELBW infants was lower than it is at present and before the use of surfactant. These youth also grew up in Canada, a developed country where medical care and education are universally available. Finally, we did not compare scale means in this study because these...
results have been reported previously (Saigal et al., 2003; Szatmari et al., 1993).

In this study, we showed for the first time that a number of the CBCL and YSR DSM-oriented scales are invariant across ELBW and NBW groups regardless of whether these are rated by parents or youth, or in childhood or adolescence. These findings also suggest that many of the previously reported group differences in psychiatric risk, particularly inattention, anxiety, and oppositional behavior between ELBW survivors and NBW controls based on the CBCL may be owing to true group differences and not differential aspects of the scales or informants. However, future work examining the risk of emotional and behavioral problems in premature offspring should examine and establish measurement invariance of the scales that they use to define psychopathology before examining group differences in these constructs.

**Supplementary Data**

Supplementary data can be found at: http://www.jpepsy.oxfordjournals.org/.

**Acknowledgments**

We thank our research assistants Susan King, Ellie Deveau, Carolyn Cuthbert, and Lorraine Hoult for assistance with data collection.

**Funding**

This work was supported by grant Number ESPM 85-201, Hospital for Sick Children Foundation, Toronto Ontario. Dr. Van Lieshout holds the Albert Einstein/Irving Zucker Chair in Neuroscience at McMaster University. Dr. Boyle holds the Canada Research Chair in the Social Determinants of Child Health. Dr. Ferro is supported by the Hamilton Health Sciences Research Early Career Award.

**Conflicts of interest:** None declared.

**References**


