Brief Report: Testing the Factorial Invariance of the CBCL Somatic Complaints Scale as a Measure of Internalizing Symptoms for Children with and without Chronic Illness

Deborah Friedman,[1] PhD, Fred B. Bryant,[2] PhD, and Grayson N. Holmbeck,[2] PhD
[1]*Bradley/Hasbro Research Center, Brown University Medical School and* [2]*Loyola University Chicago*

Objective To examine the factorial invariance of the Somatic Complaints subscale of the Child Behavior Checklist as a measure of Internalizing Behavior Problems across a sample of children with and without spina bifida. **Methods** Multisample confirmatory factor analysis was used to compare mother and father report on the Somatic Complaints subscale across a sample of children with spina bifida and a matched comparison sample of able-bodied children ages 8 through 11 years ($N = 68$ for mother report in each group; $N = 54$ for father report in the spina bifida group and 53 for the able-bodied group). **Results** Although there were no significant between-group differences in the magnitude of factor loadings, significantly more variance in scores on the Somatic Complaints scale was unrelated to Internalizing Behavior Problems for the spina bifida group, compared to the able-bodied group. There were no between-group differences when father data were analyzed, but the latent variable of Internalizing Behavior Problems explained little variance in the Somatic Complaints scale for either group. **Conclusions** Maternal report of Somatic Complaints on the CBCL does not appear to measure Internalizing Behavior Problems in the same manner across groups of children with and without spina bifida. This suggests that the Somatic Complaints subscale should be interpreted with caution when measuring Internalizing Behavior Problems within this population.

Key words internalizing; SEM; somatic complaints; spina bifida.

The Child Behavior Checklist (CBCL; Achenbach, 1991, 2001) has been widely used in research to measure behavioral problems in children who have a chronic illness, but some researchers have suggested that the CBCL should be interpreted with caution in this context (Drotar, Stein, & Perrin, 1995; Perrin, Stein, & Drotar, 1991). It has been argued that the Somatic Complaints (SC) subscale, in particular, may contain considerable bias when used with populations of children with chronic illness. Specifically, this subscale contains items that tap physical symptoms and is one of the three subscales, along with the Anxious/Depressed and Withdrawn subscales, that comprise the Internalizing Behavior Problems (IBP) composite score. While the SC subscale was intended to tap physical symptoms with no medical basis, parents of children with chronic illness may be unable to discriminate between physical symptoms that are related to an organic cause and those that may be stress-related. Scores on the CBCL SC scale may, therefore, be erroneously inflated in chronic illness groups and contribute to an overestimation of internalizing problems. Physical symptoms that reflect psychosocial difficulties may also be erroneously attributed to a child's illness, which would also contribute error variance to the IBP scale, causing it to be unreflective of true internalizing problems that children with chronic illness may be experiencing (Drotar et al., 1995; Perrin et al., 1991). These issues complicate the interpretation of the SC subscale and make its appropriateness for use with children who have a chronic medical condition questionable.

All correspondence concerning this article should be addressed to Deborah Friedman, PhD, Bradley/Hasbro Children's Research Center, The CORO Building, Suite 204, One Hoppin Street, Providence, RI 02903. E-mail: dfriedman@lifespan.org.

Journal of Pediatric Psychology 32(5) pp. 512–516, 2007
doi:10.1093/jpepsy/jsl051
Advance Access publication January 5, 2007

As a consequence, some researchers have deleted the SC subscale when interpreting internalizing scores on the CBCL (Wallander, Varni, Babani, Banis, & Wilcox, 1988; Wallander & Varni, 1989).

In one of the few studies that examined these issues directly, Holmes and colleagues (1998) deleted seven items of the SC scale that were rated by nine medical personnel as being related to the physical symptoms of diabetes. Using this revised SC subscale resulted in lower Internalizing and Total Behavior Problem scores for both children with diabetes and healthy controls, but did not eliminate group differences. While such evidence suggests that deleting the SC subscale does not eliminate differences between chronically ill and healthy children on the CBCL IBP scale, the methodology used in these studies does not adequately establish whether the SC subscale measures IBP in the same way for healthy children and children with chronic illness. In the present study, we used multisample confirmatory factor analysis (CFA) to examine whether the CBCL SC subscale measures IBP equivalently across a sample of children with spina bifida and a comparison sample of children without spina bifida (Joreskog & Sorbon, 2001; Kline, 2005). Children with spina bifida may be a particularly appropriate sample with which to test these measurement questions given that children with spina bifida are at increased risk for developing internalizing problems, and that many of the physical complaints frequently associated with spina bifida (i.e., headaches related to shunt malfunction, stomachaches related to urinary tract infections, and skin problems associated with pressure sores) directly overlap with items on the CBCL SC subscale (Appleton et al., 1997). In examining the factorial invariance of the SC scale across these samples, we addressed two basic questions: (a) whether the SC subscale is equally reflective of higher-order IBP for children with and without spina bifida; and (b) whether the amount of variability in the SC subscale that is unrelated to the higher-order construct of IBP (i.e., the unique error variance of the SC subscale) is equivalent across the two groups.

We hypothesized that both mother and father report of somatic complaints would load highly on the latent construct of IBP for the sample of children without chronic illness. It was predicted that both mother and father report of their child's somatic complaints would (a) be less reflective of internalizing problem behavior (i.e., have a lower factor loading), and (b) have more measurement error (i.e., greater unique error variance) when used as an indicator of internalizing problem behavior in a sample of children with spina bifida than for the matched sample of able-bodied youth.

Method

Participants

Participants in this study were part of a larger longitudinal investigation of psychosocial adjustment and family relationships during the transition to adolescence for children with spina bifida (e.g., Holmbeck et al., 2003). During the first data collection period (Time 1), 68 families with a child aged 8 or 9 years who had spina bifida were interviewed. The comparison sample consisted of 68 families with an able-bodied child. Data collection at Time 2 occurred 2 years after the initial interview with a retention rate of 99%. The total sample completing both Time 1 and Time 2 interviews consisted of 67 families of children with spina bifida and 66 families in the comparison group. Participants from the spina bifida and comparison group were matched on the following demographic variables: child age, child gender, child ethnicity, birth order, family structure (intact or not intact), socioeconomic status, and age of parents. Details regarding recruitment and matching of samples are provided in Holmbeck et al. (2003).

Measures

The CBCL (Achenbach, 1991) is a 116-item measure assessing internalizing and externalizing symptoms, as well as three areas of competence (activities, social, and school). The subscales that combine to form the internalizing problem scale are Withdrawn, Anxiety/Depression, and Somatic Complaints. In the present analyses, we analyzed maternal and paternal reports on the CBCL separately, and we averaged T1 and T2 data to create more reliable subscales. In the analysis of mother report, $N = 68$ for each group (spina bifida and able-bodied). In the analysis of father report, $N = 53$ for the able-bodied group and $N = 54$ for the spina bifida group. Although we report the results for data averaged across the two time points separately for mothers and fathers, analyses of mother and father data separately at Times 1 and 2 revealed a nearly identical pattern of results at each time point.

Procedure

Families were interviewed at Time 1 when children were aged 8–9 years and again at Time 2, 2 years later. Data at both Time 1 and 2 were collected by trained graduate and undergraduate psychology students during a home visit. During both sessions, the family completed questionnaire

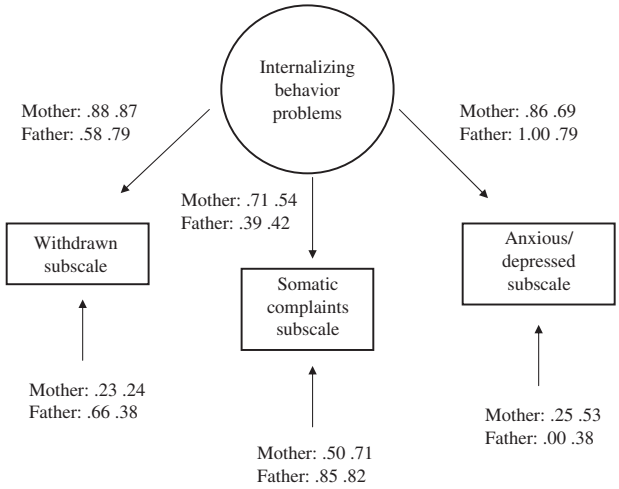

Figure 1. Completely standardized parameter estimates for the one-factor CFA model imposed on mother and father reports for the able-bodied group ($N = 68$) and spina bifida group ($N = 68$). Parameter estimates for the able-bodied sample are in regular font; parameter estimates for the spina bifida sample are in bold font.

packets and participated in a series of videotaped family interaction tasks.

Results

Analytic Strategy

We used CFA via LISREL 8 (Joreskog & Sorbom, 2001) to examine the parameter estimates (i.e., factor loading and unique error variance) of the SC subscale and to compare estimates across the spina bifida and able-bodied samples separately for mother and father reports. CFA allows one to test hypotheses about equality in parameter estimates for a particular model both within a single group and across multiple groups (Joreskog & Sorbom, 2001).

The first step in comparing parameter estimates across the spina bifida and able-bodied groups was to establish a baseline measurement model of IBP for each group separately. Imposing the model separately on each group allowed us to create baseline models to which nested models with equality constraints could be directly contrasted. These baseline models included three measured variables: parent report on each of the following scales, Anxious/Depressed, Withdrawn, and SC. Due to sample size constraints, we used the three CBCL subscale scores rather than their individual constituent items as measured variables thus making the CFA baseline model a second-order measurement model (Bagozzi & Edwards, 1998).

After the baseline measurement model was established, we then conducted a multisample CFA in which the model was respecified with equality constraints, first on the factor loading of the SC subscale, and then on both the factor loading and unique error variance of the SC subscale. In other words, these parameters were constrained to be equal across the samples of children with and without spina bifida, and then models with and without equality constraints were compared. We used the maximum-likelihood chi-square value to assess model fit, so that a nonsignificant difference in chi-square values when contrasting the models with and without equality constraints signifies between-group equivalence of the particular parameters in question. If the chi-square value for the model with equality constraints is significantly larger than the chi-square value for the baseline model, then this result signifies a between-group difference in the particular parameters in question (Joreskog & Sorbom, 2001).

Establishing a Measurement Model for the CBCL

Figure 1 provides a schematic representation (including completely standardized parameter estimates for the loadings and error terms of each CBCL subscale) of the baseline measurement model imposed first on mothers' data separately for the able-bodied and spina bifida groups and then on father's data for the able-bodied and spina bifida groups. When imposed on the data of the able-bodied sample using mother report, the completely standardized factor loadings for the one-factor CFA model were relatively strong for all three subscales: Withdrawn (.88), Anxious/Depressed (.86), and SC (.71). Consistent with our hypothesis, completely standardized factor loadings for mother report in the spina bifida sample were relatively strong for the Withdrawn (.87) and Anxious/Depressed (.69) subscales, but lower for the SC subscale (.54).

When the model was imposed on the able-bodied group using father report on the three subscales, the model would not converge to provide parameter estimates. The intermediate solution revealed that the estimate of unique error variance for the Anxious/Depressed subscale was negative, a mathematical impossibility known as a ''Heywood case'' (Wothke, 1993). To overcome this technical problem, we constrained the unique error variance of the Anxious/Depressed subscale to be non-negative (Joreskog & Sorbom, 2001; Wothke, 1993). The model then converged and yielded the following results. The completely standardized estimates for the scales were moderate for the Withdrawn subscale (.58), high for the Anxious/Depressed subscale (1.00), and low for the SC subscale (.39). Findings were similar

for father report in the spina bifida sample, with strong factor loadings for the Withdrawn (.79) and Anxious/Depressed (.79) subscales, and a lower loading for the SC scale (.42).

Assessing the Factorial Invariance of the Scale across Groups

Contrary to our hypothesis, when contrasted with the baseline CFA models for each group, the first multisample CFA model including mother-reported data indicated that the factor loading of SC was invariant across the two groups, $\Delta\chi^2(1) = 0.02$, $p < .89$. Confirming our second hypothesis, when contrasted with this first multisample model, the second multisample CFA model revealed that the unique error variance of SC was significantly larger when mothers of children with spina bifida were the reporters ($\delta = 4.51$) than when mothers of able-bodied children were the reporters ($\delta = 1.81$), $\Delta\chi^2(1) = 10.59$, $p < .002$. Viewed from the perspective of classical test theory (McDonald, 1999), the pattern of results for mother-report of the Internalizing factor indicates that the true score component (i.e., factor loading) of SC is invariant for the spina bifida and the able-bodied groups, but that the error component (i.e., unique error variance) is larger for the spina bifida group.

When the same test of factorial invariance was conducted using father-reported data, both the factor loading, $\Delta\chi^2(1) = 1.73$, $p < .19$, and unique error variance, $\Delta\chi^2(1) = 0.38$, $p < .54$, of SC were invariant across the two groups. Thus, in contrast to mother-reported data, fathers of children with spina bifida and fathers of able-bodied children had equivalent factor loadings and equivalent unique error variances for SC in the one-factor CFA model.

Discussion

The present study examined the validity of the CBCL SC scale as a measure of IBP when used with a population of children who have a chronic medical condition. Multisample confirmatory factor analysis revealed that although the SC subscale had an equivalent loading on the Internalizing factor for both groups of children with and without spina bifida, *mother* report of SC contained significantly more measurement error when this report was used as an indicator of IBP in the spina bifida group compared to the able-bodied group. Given the findings for the mother-reported data, we conclude that the SC scale contains more variance unrelated to IBP when used with the spina bifida sample. This unique error variance represents the sum of both random (error) and specific (constructs other than IBP) influences on responses (Bollen, 1989; Kline, 2005).

Several researchers have suggested that the SC scale lacks construct validity when used with children who have a chronic medical condition due to the real physical symptoms that these children face. When a parent of a child with a chronic illness reports on the child's somatic complaints, real physical symptoms may be confused with symptomatology reflecting psychosocial disturbance (Drotar et al., 1995; Perrin et al., 1991). The results of the present study provide evidence to support this theory.

Results of multisample confirmatory factor analyses examining the validity of *father* reports of SC on the CBCL yielded somewhat different findings. There were no significant differences in factor loadings or unique error variances for the SC subscale as a measure of IBP when comparing father reports in the spina bifida and able-bodied samples. Father report of SC appeared to be a relatively unreliable indicator of IBP across both groups.

Many interpretations may be provided to explain the difference in the utility of father and mother report of the child's somatic complaints within the able-bodied sample. One explanation is that if the mother tends to be the primary caregiver in the household, she may spend more time with the child and therefore, be more likely to have knowledge about the child's physical symptoms. Another possibility is that mothers may be more empathic than fathers with regard to their child's level of stress or internalizing behavior problems, and may therefore be more likely to ask their child about his or her feelings and physical symptoms.

There are several limitations to this study. First, the sample sizes were small, particularly for this type of statistical analyses (CFA), and replication is needed before these results can be confidently generalized to the larger population of children with spina bifida. Second, the results of our analyses may not apply to children with different types of chronic illness conditions. Third, given that some changes have been made to the three subscales that comprise the IBP scale in a revised version of the CBCL, results of this study need to be replicated with the revised version to confirm generalizability across these measures (Achenbach & Rescorla, 2001). In addition, though T1 and T2 data were used to test the factorial invariance of the SC subscale across groups, changes in medical status that may have occurred over the two time points and that may affect parent report were not measured. Finally, our samples included only

8–11 year-old children; thus, results may not apply to younger or older children.

Despite these limitations, this study provides an example of a useful statistical methodology for examining specific questions related to construct bias when measuring an attribute across multiple groups. An advantage to this method is that is allows one to examine whether the amount of variability both related (true score) and unrelated to the higher-order construct (the measurement error) is captured in an equivalent manner across groups. This method typically requires a greater sample size, however, to examine measurement models larger than the three factor second-order model investigated in the present study.

The present results provide evidence about the measurement of internalizing symptoms using the CBCL with children who are chronically ill. Our findings suggest that maternal reports of SC also reflect children's medical concerns in addition to internalizing symptoms with children who have spina bifida. Therefore, caution should be exercised when selecting an instrument to measure internalizing symptoms in pediatric research, because biased measurement may lead to misinterpretation of group differences. In addition, detailed attention to the issue of somatic complaints within the clinical interview, as well the administration of alternative measures of internalizing symptoms that are not based on somatic complaints (e.g., the CDI), may be necessary within the framework of a psychological assessment of a child with chronic illness.

Conflict of interest: None declared.

Received March 16, 2006; revisions received September 15, 2006 and December 2, 2006; accepted December 2, 2006

References

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18, YSR, and TRF profiles*. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Appleton, P. L., Ellis, N. C., Minchom, P. E., Lawson, V., Boll, V., & Jones, P. (1997). Depressive symptoms and self-concept in young people with spina bifida. *Journal of Pediatric Psychology, 22(5):707–722.*

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45–87.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the child behavior checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology, 24*, 184–192.

Holmbeck, G. N., Westhoven, V. C., Shapera Phillips, W., Bowers, R., Gruse, C., Nikolopoulos, T., et al. (2003). A multimethod, multi-informant, and multidimensional perspective on psychosocial adjustment in preadolescents with spina bifida. *Journal of Consulting & Clinical Psychology, 71*, 782–796.

Holmes, C. S., Respess, D., Greer, T., & Frentz, J. (1998). Behavior problems in children with diabetes: Disentangling possible scoring confounds on the child behavior checklist. *Journal of Pediatric Psychology, 23*, 179–185.

Joreskog, K. G., & Sorbom, D. (2001). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* 2nd ed. New York: Guilford.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Perrin, E. C., Stein, R. E. K., & Drotar, D. (1991). Cautions in using the child behavior checklist: Observations based on research about children with a chronic illness. *Journal of Pediatric Psychology, 16*, 411–421.

Wallander, J. L., & Varni, J. W. (1989). Disability parameters, chronic strain, and adaptation of physically handicapped children and their mothers. *Journal of Pediatric Psychology, 14*, 23–42.

Wallander, J. L., Varni, J. W., Babani, L., Banis, H. T., & Wilcox, K. T. (1988). Children with chronic physical disorders: Maternal reports of their psychological adjustment. *Journal of Pediatric Psychology, 13*, 197–212.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen, & J. S. Long (Eds.). *Testing structural equation models* (pp. 256–293). Newbury Park CA: Sage.