THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY
ACAMH
Journal of Child Psychology and Psychiatry 51:9 (2010), pp 1040–1049 doi:10.1111/j.1469-7610.2010.02244.x

The comparability of mother-report structured interviews and checklists for the quantification of youth externalizing symptoms

Melanie A. Dirks[1] and Michael H. Boyle[2]

[1]McGill University, Montreal, Canada; [2]Offord Centre for Child Studies, McMaster University, Canada

Background: Although structured interviews are assumed to be scientifically superior to checklists for measuring youth psychopathology, few studies have tested this hypothesis. Interviews place a much greater burden on respondents, making it critical to determine their added value when quantifying psychiatric symptoms. **Methods:** Confirmatory factor analysis was used to compare interviews and checklists in community (N = 251) and clinically referred (N = 406) samples of youth aged 5 to 17 years. We examined the associations between mother-reported externalizing symptoms assessed by interview versus checklist against (a) teacher-reported externalizing symptoms, and (b) child's gender, academic performance, single- versus two-parent family, and family income. Models in which associations were estimated freely were contrasted to models in which the interview and the checklist were constrained to have equal associations with the variables. Finding these models fit comparably would suggest no difference between interviews and checklists. **Results:** In the community sample, both the constrained and unconstrained models provided comparable fit to the data, suggesting no marked differences between interviews and checklists. In the clinical sample, associations with the interview were generally stronger. Reducing the number of items on the interview to match those on the 6-item checklist eliminated these differences, suggesting that the increased reliability of the interview scales, afforded by additional items, enhanced their quantification of psychopathology. **Conclusions:** Consistent with previous studies, interviews were not notably superior to checklists for the measurement of externalizing symptoms. When only a few items are used, small performance differences between checklists and interviews may be due to scale length. **Keywords:** Measurement, externalizing symptoms, interview, checklist.

We need valid, reliable measures of children's psychological symptoms for both clinical practice and research endeavors. Clinically, such assessments provide information valuable for planning interventions, as well as tracking progress in treatment. These measures also facilitate important research such as identifying the associated features of psychiatric conditions. In addition to accuracy, these measures must be as efficient as possible. The burden of lengthy measures reduces the number of assessments that can be administered in research studies and may increase levels of missing data. In the clinic, excessively long assessments may be frustrating for families and expensive to administer.

The goal of this study is to determine if checklists and interviews provide comparable quantification of youth psychiatric symptoms. To achieve this goal, we used two methods to examine the equivalence of these measurement strategies in both community and clinically referred samples. First, we compared the associations between externalizing symptoms assessed by mother report using an interview or a checklist and the same symptoms reported by teachers. Second, we examined the strength of association between each of the assessment techniques and constructs with established associations to externalizing symptoms (i.e., gender, academic performance, single-parent status and family income). Finding comparable associations between these variables and symptoms identified by interviews versus checklists would suggest that these assessment approaches are equivalent.

Checklists and structured interviews are the two primary methods of assessing child psychopathology (Roberts et al., 1998). Interviews can be either respondent or interviewer based (Shaffer, Fisher, & Lucas, 1999). Respondent-based interviews feature highly structured questions, giving the interviewer little latitude to interpret these queries or informants' responses. Interviewer-based interviews provide a much greater degree of flexibility in both the definition of symptoms and the questions asked (Angold & Fisher, 1999). This study focused on respondent-based interviews as this approach is similar to checklists in their heavy reliance on respondent judgment.

It is generally assumed that researchers and clinicians are forced to accept a trade-off between the scientific superiority of the interview and the practical advantages of the checklist; however, few studies have tested this comparison explicitly. Several features of interviews are likely to yield greater reliability and validity. First, interviews usually include more items, which enhances the coverage of

Conflict of interest statement: No conflicts declared.

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

constructs. These additional items should also increase reliable variance of measurement, which will improve statistical power by reducing the attenuating effects of measurement error (Boyle & Pickles, 1998). Furthermore, interview administration is highly standardized: the presence of an interviewer eliminates concerns about informants' ability to read and provides respondents with opportunities to clarify questions. In contrast, checklists are comprised of fewer items, take little time to complete, pose minimal burden and can be administered under a variety of different conditions (e.g., by mail).

Only a handful of studies have compared the measurement properties of checklists and interviews. Lack of a 'gold standard' constitutes an important methodological challenge to this comparison: unlike many physical disorders, there are no markers that identify definitively the presence of childhood emotional or behavioral disorders, and researchers must use a variety of methods to ascertain the equivalence of different assessment techniques.

One common approach to assessing measurement equivalence is to compare the strength of association between psychopathology assessed with an interview or a checklist and external 'validator' variables – constructs with established associations to children's psychopathology (Jensen et al., 1996). Two studies have shown that symptoms assessed with the Diagnostic Interview Schedule for Children (DISC-IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) and the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) exhibited comparable associations with several constructs including school dysfunction, use of mental health services, and family functioning (Gould, Bird, & Staghezza, 1993; Jensen et al., 1996). Furthermore, over a three-year period, predictive associations were comparable between these same two approaches and similar endpoints (Ferdinand et al., 2004).

Although small in number, studies examining comparability between checklists and interviews point to equivalence. We aimed to contribute to this evidence base by assessing the comparability of mother-reported youth externalizing symptoms (i.e., attention deficit hyperactivity disorder, ADHD; conduct disorder, CD; and oppositional defiant disorder, ODD) obtained through respondent-based interviews versus checklists implemented in both community- and clinic-based samples. These analyses contribute to the existing literature in two ways. One, the study included general population and clinical samples. Previous studies have been limited to specific populations, such as Puerto Rican children (Gould et al., 1993) or military families (Jensen et al., 1996), and it is unclear whether these findings will generalize to other community samples. Moreover, only one study has examined the comparability of interviews and checklists in clinical populations, and the sample was quite small ($N = 96$; Ferdinand et al., 2004). Several features of clinically referred samples may affect the equivalence of measures, such as increased levels of symptomatology and higher levels of co-morbidity (Gould et al., 1993). Two, in the clinic sample we examined a much briefer checklist. Previous work has focused on the CBCL (e.g., Ferdinand et al., 2004; Gould et al., 1993; Jensen et al., 1996). CBCL scales, although shorter than interviews, are still quite lengthy. In the current work, we examined the comparability of the Brief Child and Family Phone Interview (BCFPI; Boyle et al., 2009; Cunningham, Boyle, Hong, Pettingill, & Bohaychuk, 2009), a screening instrument that has only six items per scale.

We used two approaches to assess measurement equivalence. First, we examined the associations between measures of psychopathology derived from interviews versus checklists completed by mothers and teacher-reported externalizing symptoms. Second, we assessed the associations between interviews versus checklists and four external variables: gender, academic performance, single-parent status and family income. These variables were selected because they have established associations with externalizing symptoms. Males display higher rates of externalizing problems (Rey, Walter, & Soutullo, 2007; Spetie & Arnold, 2007). Economic disadvantage (e.g., Costello et al., 1996), living in a single-parent family (Duncan, Brooks-Gunn, & Klebanov, 1994; Florsheim, Tolan, & Gorman-Smith, 1998), and poorer academic performance (see Bradshaw, Buckley, & Ialongo, 2008) are all associated with greater behavior difficulties. Importantly, measurement of these constructs was 'independent' of psychopathology in the sense that it did not rely on behavioral reports from mothers.

Method

Participants

The *community sample* was 251 youth aged 6 to 17 years ($M = 11.59$, $SD = 3.39$, 51% male). Participants were a stratified random sample (76.3% response) drawn from a simple random sample of 1,751 youth (78.9% response) taken in 1989 from those attending all public schools in an industrialized, urban setting (94 schools and 27,629 students; Boyle et al., 1993a, 1993b). The strata consisted of 8 mutually exclusive groupings derived from age in years (6 to 11, 12 to 16), sex and high versus low symptom scores based on the original Ontario Child Health Scales (Boyle et al., 1987). The sampling weights were the inverse probability of selection within each stratum.

The *clinical sample* was families with youth aged 5 to 17 years seeking mental health services from community providers in three Ontario cities between September 2003 and December 2004. Families had to reside within the site catchment areas and be English speaking. Youth with clinical evidence of a neurological

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

disorder such as epilepsy, a serious medical problem, psychotic symptoms or suicidal behavior were excluded. The final sample consisted of 406 youth (M = 10.17 years of age, SD = 2.97), which represented about 19% of families referred during the recruitment period. Reasons for this low recruitment have been documented elsewhere (Boyle et al., 2009), and include (a) high burden of participation (five separate measurement occasions); (b) enlistment occurring as families learned that they would be placed on a four- to six-month waitlist to receive services; and (c) families not receiving compensation for their time. Levels of psychopathology and family functioning were similar for participants and non-participants. Younger children and intact families were overrepresented among participants.

Measures

Ontario Child Health Study-Revised (OCHS-R) Scales. The OCHS-R scales were adapted from the original OCHS scales (Boyle et al., 1987) to assess symptom criteria specified in DSM-III-R. We focused on the ADHD, CD, and ODD scales. Each item is scored 0 (*never or not true*), 1 (*sometimes or somewhat true*), or 2 (*often or very true*). Raw scores are added together to form a scale score for each disorder. Parent and teacher versions of the OCHS-R scales are almost identical. Detailed information on scale development, reliability, and validity is available (Boyle et al., 1993a). In the current study, Cronbach's alphas for mother-reported symptoms were: ADHD, .91; CD, .70; ODD, .87.

Diagnostic Interview for Children and Adolescents-Revised (DICA-R). The DICA-R (Reich & Welner, 1988) is a lay-administered structured interview. The interviewer training procedures, modifications to the interview itself, and psychometric properties of the DICA-R in our study appear elsewhere (Boyle et al., 1993b). To facilitate comparison to the checklist, we computed scale scores for the externalizing behavior modules. Participants' responses were coded as 0 (*no*), 1 (*sometimes*), and 2 (*yes*). Sums were computed for ADHD, CD, and ODD. Cronbach's alphas were: ADHD, .87; CD, .79; ODD, .79.

Brief Child and Family Phone Interview (BCFPI). The BCFPI is a parent-report measure of emotional and behavioral problems of children aged 3 to 18 years referred for mental health services. Three six-item scales corresponding to the DSM categories of ADHD, CD, and ODD were used in this study. Taken from the OCHS-R scales, the BCFPI items are scored 0 (*never true*), 1 (*sometimes true*), or 2 (*often true*) and summed within each DSM category to obtain scale scores from 0 to 12. Detailed information on measurement development and psychometric properties is available (Boyle et al., 2009; Cunningham et al., 2009). In our sample, Cronbach's alphas were: ADHD, .86; CD, .68; ODD, .83.

Diagnostic Interview Schedule for Children-IV (DISC-IV). The DISC-IV is a computer-assisted, structured interview that can be administered by non-clinicians (Shaffer et al., 2000). Interviewers received 1.5 days of training on-site from an expert trainer seconded from Columbia University. ADHD, CD and ODD were among the modules completed by mothers. Scale scores were computed by summing responses coded as 0 (*no*), 1 (*sometimes*), and 2 (*yes*) for each condition. Cronbach's alphas were: ADHD, .90; CD, .82; ODD, .82. The DISC-IV scales consist of many more items than the comparison checklist (the BCFPI). For this reason, we also calculated reduced DISC-IV scales by reviewing the items on the BCFPI, identifying the most similar items on the interview and dropping remaining items. Cronbach's alphas for the reduced ADHD, CD, and ODD scales were .75, .62, and .72, respectively.

Gender, academic performance, single-parent status and family income. Measures of these constructs were included as external 'validator' variables. In both the community and clinic samples, academic performance was assessed using a teacher rating in response to the statement 'this student's current academic achievement across all areas' from 1 (*near the bottom of the class*) to 5 (*near the top of the class*), and single-parent status was assessed with the question 'Are you a single parent or do you live with a spouse or partner?' In the community sample, family income was assessed using mother-reported total household income from 1 (*less than $30,000*) to 8 (*more than $80,000*); in the clinic sample, mothers reported their household income in response to the question 'What is the best estimate of your total household income from all sources in the last tax year?'

Procedures

Written, informed consent was required from mothers and adolescents aged 12 to 17 years; verbal assent was required from 5- to 11-year-olds. All study procedures, including provisions for obtaining consent and safeguarding privacy, were approved by the Research Ethics Board at McMaster University.

Community sample. Children were assessed three times: baseline, 6 weeks and 8 weeks. At baseline, mothers (during a home interview) and teachers (by mail) completed the revised OCHS-R scales (Boyle et al., 1993a). At 6 weeks, mothers and teachers completed the OCHS-R scales a second time, and each mother was interviewed using the DICA-R. The administrations of the OCHS-R scales and the DICA-R were done independently, several days apart, and their order of completion randomized. At 8 weeks, the DICA-R was administered a second time to mothers.

At the scale level, the amount of missing data was quite low, with a maximum of 10% for teacher-reported CD. In addition, 7% and 19% of the data were missing from family income and teacher-report of school performance, respectively. In order to use sampling weights in analyses conducted with AMOS, it is necessary to first construct a covariance matrix in SPSS. In doing so, any cases with missing data are eliminated. To prevent this excessive sample loss, missed responses were given estimated values using the imputation procedure in the program WinMice (Jacobusse, 2005). This

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

program draws imputations from the multivariate distribution, estimated from the incomplete data in a Gibbs sampling process. The steps involve selecting a variable with missing values, specifying an imputation method depending on the measurement level (e.g., linear-regression model for a quantitative measure) and naming predictor variables. Model parameters are derived from multiple iterations, and imputations for the missing values are based on parameter estimates from the very last iteration of the Gibbs sampler. Separate models were derived for each variable with missing information and run simultaneously. All of the variables used in the analyses served as predictor variables in each model. Five iterations of each model were run and results were combined by WinMice into a final analytic dataset. Data were presumed missing at random and adequately modeled by the predictor variables. We believe that departures from this assumption would have only a minor impact on estimates and standard errors (Collins, Schafer, & Kam, 2001). Distributions of all variables were checked and, when necessary, appropriate transformations were applied to reduce skewness and kurtosis.

Clinic sample. There were five measurement occasions: baseline, 1 month, 2 months, 12 months and 13 months. At baseline, during an intake phone interview, mothers answered screening questions on youth problem behavior included in the BCFPI (Boyle et al., 2009). At 1 month, each mother completed the DISC-IV during a home interview. At 2 months, mothers were interviewed by phone a second time using the BCFPI. At 12 and 13 months, mothers were re-administered the DISC-IV (home interview) and BCFPI (telephone interview), in that order. Less than 2% of the data was missing from any of the mother-reported scales. The OCHS-R scales were mailed to teachers for completion at 1 month and 12 months. Eighty-seven teachers (21%) did not return their questionnaires. On the completed questionnaires, several items (e.g., has broken into someone else's home, building, or car) were skipped by the majority of teachers, probably for lack of knowledge. A procedure identical to that described previously was used to address missing data at the item level for all teacher-rated symptoms of ADHD, CD and ODD, regardless of how many items the teacher had completed. Distributions of all variables were checked and, when necessary, appropriate transformations were applied to reduce skewness and kurtosis.

Analyses

Within each sample, two sets of analyses were conducted. First, we examined the association between teacher- and mother-reported externalizing symptoms (see Figure 1). Externalizing symptoms were represented as three latent variables (i.e., interview, checklist, teacher), with three indicator variables (ADHD, CD, ODD) loading on each. This model was fit twice. In the first model, all parameters were estimated freely. In the second model, the path coefficients linking the latent constructs representing the interview and the checklist to teacher-reported externalizing symptoms were constrained to be equal, as were the covariances between

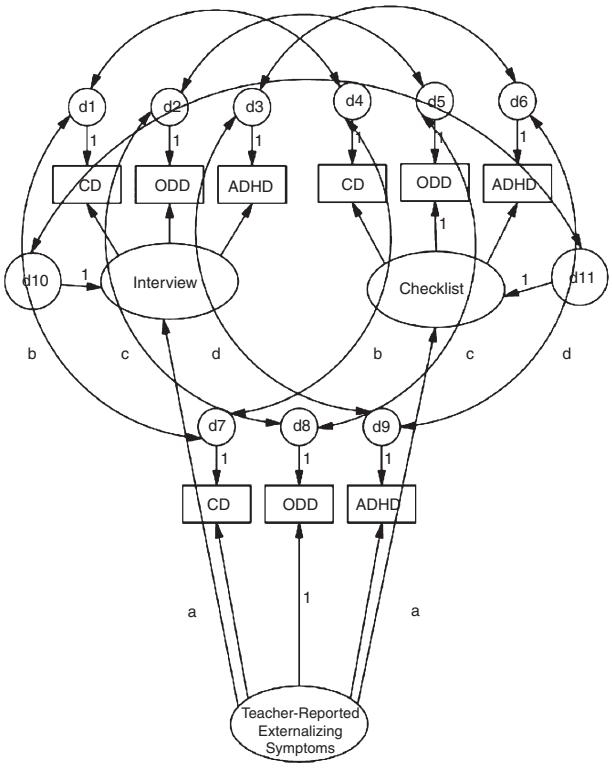

Figure 1 Model linking mother- and teacher-reported externalizing symptoms. *Note.* Model was calculated twice. In the first model, all parameters were estimated freely. In the second, all paths sharing a letter were constrained to be equal. ADHD = attention deficit hyperactivity disorder. CD = conduct disorder. ODD = oppositional defiant disorder

the residual terms for ADHD, CD, and ODD assessed by each method. Second, we examined the associations between the external variables – gender, academic performance, single-parent family status and family income – and mother-reported externalizing symptoms assessed by both interview and checklist (see Figure 2).

We considered four questions in evaluating the equivalence of interviews and checklists. One, is there a significant change in model fit when parameters are constrained? Finding that the constrained models fit the data as well as the unconstrained models would suggest that the associations between symptoms assessed by interview and the predictor variables were equivalent to the associations between checklist-reported symptoms and the predictor variables. To assess comparability, we computed the likelihood ratio test derived from loss of fit between two nested models. A significant χ^2 value indicates that adding parameter constraints resulted in a poorer fitting model. This test is sensitive to sample size and, with large samples, even small differences will be significant (Kline, 2005). For this reason, we examined two alternate fit indices: the CFI and RMSEA. Recently, it has been suggested that differences in CFI between models less than .01 indicate equivalence (Vandenberg & Lance, 2000). Two, does the constrained model provide adequate fit to the data? Finding that the constrained model fit the data adequately (i.e., CFI exceeding .95 and RMSEA with a lower-bound confidence interval overlapping .06; Hu &

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

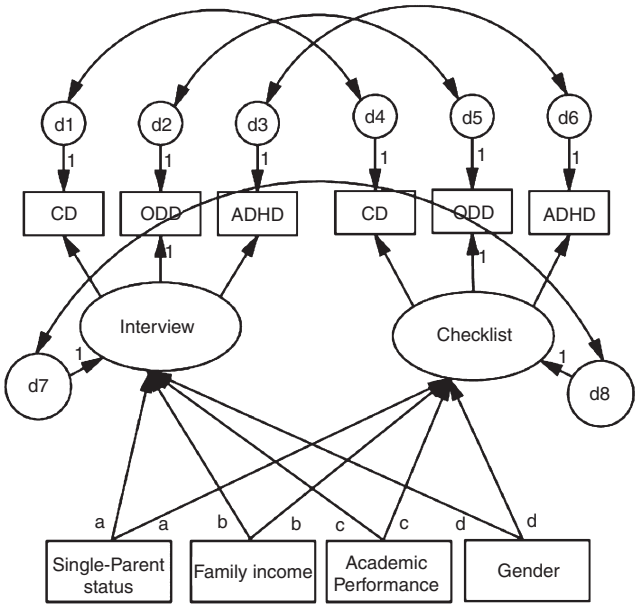

Figure 2 Model linking mother-reported externalizing symptoms and external 'validator' variables. *Note.* Model was calculated twice. In the first model, all parameters were estimated freely. In the second, all paths sharing a letter were constrained to be equal. ADHD = attention deficit hyperactivity disorder. CD = conduct disorder. ODD = oppositional defiant disorder

Bentler, 1999) would suggest that differences between the interview and the checklist are modest; if the associations were vastly different, constraining them to be equal should result in a model that does not fit. Three, does the pattern of associations consistently favor either the interview or checklist? It is possible that measurement-related error or sampling variability may have resulted in a given association being stronger. In this case, constraining the two models to be equal may result in worse fit, but this may not reflect the superiority of the measure. Four, are the results consistent across tests (i.e., associations with teacher-assessed symptoms and external validator variables)?

Results

Mother- and teacher-reported externalizing symptoms

Community sample. Zero-order correlations between all measured variables are presented in the appendix. We began by fitting the unconstrained model. For the unconstrained model, the chi-square test was significant, $\chi^2(15) = 29.82$, $p < .05$, suggesting sub-optimal fit to the data. However, the CFI (.99) and RMSEA (.064; 90% CI = .029–.097) were acceptable. Standardized path coefficients and correlations are presented in Table 1. Externalizing symptoms, as assessed by both the interview and checklist, were associated with teacher-reported symptoms, $\beta = -.51$ $p < .01$; $\beta = -.45$, $p < .01$; respectively. (These associations are negative because an inverse transformation was performed on each of the teacher-reported symptom scales.) Next, we fit the constrained model. The likelihood ratio test was not significant, $\chi^2(4) = 7.44$, $p > .05$, suggesting that constraining the relationships

Table 1 Associations between teacher-reported externalizing symptoms and mother-reported externalizing symptoms using an interview or a checklist

	Unconstrained model Interview/Checklist	Constrained model Interview/Checklist
Community sample		
Externalizing symptoms[a]	−.51**/−.45**	−.45**/−.48**
Residual correlations		
ADHD	−.24**/−.30**	−.23**/−.30**
CD	−.07/.09	.03/.06
ODD	.26/.18	.14/.17
Clinic sample – full interview		
Externalizing symptoms	.34**/.26**	.30**/.30**
Residual correlations		
ADHD	.41**/.32**	.19**/.31**
CD	.35**/.19**	.35**/.18**
ODD	−.43**/−.02	−.16/−.14
Clinic sample – reduced interview		
Externalizing symptoms	.23**/.26**	.24**/.24**
Residual correlations		
ADHD	.32**/.31**	.28**/.33**
CD	.31**/.19**	.28**/.23**
ODD	−.13/−.01	−.08/−.08

Notes: CD = conduct disorder. ODD = oppositional defiant disorder. ADHD = attention deficit hyperactivity disorder. Note that constraining parameters to be equal equates the unstandardized coefficients. Standardized coefficients may still differ, given differences in standard deviations.

[a]These associations are negative because in the community sample an inverse transformation was applied to the variables for teacher-reported CD, ODD, and ADHD to reduce the skewness and kurtosis of the distributions.

*$p < .05$; **$p < .01$.

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

between interview- and checklist-assessed externalizing symptoms and teacher-reported symptoms did not result in loss of model fit. The CFI was approximately the same as the unconstrained model (.99). The RMSEA, .063, 90% CI = .032–.092, was also acceptable.

Clinic sample. This model provided good fit to the data, $\chi^2(15) = 18.15$, $p > .05$, CFI = 1.00, RMSEA = .023 (90% CI = .000–.055). Standardized path coefficients and correlations are presented in Table 1. Teacher-reported symptoms predicted significantly mother-reported symptoms as captured by both interview, $\beta = .34$, $p < .01$; and checklist, $\beta = .26$, $p < .01$. The likelihood ratio test revealed that constraining parameters resulted in markedly worse fit; $\chi^2(4) = 39.23$, $p < .01$. The CFI (.98) indicated adequate model fit; however, it was more than .01 lower than the CFI for the constrained model, again suggesting the models were not equivalent. The RMSEA (.080, 90% CI = .060–.100) indicated adequate fit. (The chi-square test for the constrained model was: $\chi^2(19) = 57.38$, $p < .01$.)

Externalizing symptoms and external variables

Community sample. Again, we began by fitting an unconstrained model in which all parameters were estimated freely. Model fit was acceptable: $\chi^2(21) = 58.66$, $p < .05$; CFI = .96; RMSEA = .086 (90% CI = .060–.112). Standardized path coefficients linking predictors and externalizing symptoms are presented in Table 2. A significant negative relationship existed between family income and externalizing symptoms as measured by interview and checklist, and between academic performance and externalizing symptoms as measured by interview and checklist. Gender shared a significant association with checklist-measured symptoms. Single-parent status was not associated with externalizing symptoms. The regression coefficients exhibited no consistent pattern of association with either the interview or the checklist. In the second model, the pathways linking each predictor to the two latent variables were constrained to be equal. The likelihood ratio test indicated that this model did not fit as well as the unconstrained model, $\chi^2(4) = 13.14$, $p < .05$. The chi-square model test was significant, $\chi^2(25) = 71.80$, $p < .01$. The CFI (.96) indicated adequate fit to the data, and was less than .01 lower than the CFI of the unconstrained model, suggesting model equivalence. The RMSEA (.088, 90% CI = .064–.112) exceeded slightly the cutoff identified by Hu and Bentler (1999); however, it does not differ statistically from the RMSEA of the unconstrained model and is still within acceptable bounds using other published criteria (e.g., Browne & Cudeck, 1993; MacCallum, Browne, & Sugara, 1996).

Clinic sample. The chi-square test for the unconstrained model was significant, $\chi^2(21) = 58.56$, $p < .01$; however, both the CFI (.97) and the RMSEA (.066; 90% CI = .047–.083) suggested adequate fit to the data. Standardized path coefficients linking predictors and externalizing symptoms are presented in Table 2. Academic performance was negatively associated with externalizing symptoms, as measured by both the interview and the checklist, and the association with the interview was stronger. Single-parent status was associated with externalizing symptoms as measured by the checklist. The likelihood ratio test indicated that there was significant loss of fit in the constrained model, $\chi^2(4) =$

Table 2 Associations between gender, academic performance, single-parent status, and family income and mother-reported externalizing symptoms using an interview or a checklist

	Unconstrained model Interview/Checklist	Constrained model Interview/Checklist
Community sample		
Gender	.03/−.15*	−.11/−.12
Academic performance	−.32**/−.21**	−.22**/−.25**
Single-parent status	.01/−.05	−.03/−.03
Family income	−.32**/−.26**	−.22**/−.24**
Clinic sample – full interview		
Gender	−.09/−.03	−.05/−.05
Academic performance	−.30**/−.24**	−.25**/−.27**
Single-parent status	−.07/−.18**	−.12/−.13
Family income	−.02/−.03	−.02/−.02
Clinic sample – reduced interview		
Gender	−.08/−.04	−.06/−.06
Academic performance	−.33**/−.26**	−.30**/−.28**
Single-parent status	−.08/−.16*	−.13/−.12
Family income	−.05/−.04	−.04/−.04

Notes: Gender dummy coded males = 0. Single-parent status dummy coded single-parent family = 0. Note that constraining parameters to be equal equates the unstandardized coefficients. Standardized coefficients may still differ, given differences in standard deviations.
*$p < .05$; **$p < .01$.

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

13.69, $p < .05$; the CFI (.96) was within .01 of the unconstrained model, suggesting model equivalence. The RMSEA (.068, 90% CI = .050–.087) indicated adequate fit. (The chi square test was significant, $\chi^2(25) = 72.26$, $p < .01$.)

Reduced interviews

In the clinic sample, the interview exhibited generally stronger associations with the predictors, a pattern not evident in the community sample. This finding could mean that the interview is more 'accurate' in a clinic sample, where levels of symptomatology and co-morbidity are higher. Alternatively, it might be due to excessive measurement error in the BCFPI scales arising from insufficient items (i.e., six items per scale). To examine the impact of scale length on measurement equivalence in the clinic sample, we reduced the number of items on the DISC-IV scales so that each was the same length as the corresponding BCFPI scale.

The two models were re-run using the reduced DISC-IV scales as indicator variables. For the model comparing mother- and teacher-reported externalizing symptoms, the unconstrained model fit the data well, $\chi^2(15) = 23.08$, $p > .05$; CFI = 1.00; RMSEA = .036 (90% CI = .000–.064; see Table 1 for path coefficients and covariances). The likelihood ratio test indicated that constraining the model did not reduce model fit significantly, $\chi^2(4) = 6.04$, $p > .05$; and the CFI remained approximately the same (.99), also suggesting model equivalence. The RMSEA, .036, 90% CI = .000–.061, indicated adequate fit, as did the chi-square test, $\chi^2(19) = 29.12$, $p > .05$. For the model comparing the associations between the interview and checklist and the external validator variables, the chi-square test for the unconstrained model was significant, $\chi^2(21) = 60.45$, $p < .01$, although the model provided adequate fit, CFI = .96; RMSEA = .068 (90% CI = .048–.089). The pattern of associations between externalizing symptoms and the external variables was unchanged and is presented in Table 2. Constraining parameters did not result in a significant loss of model fit, $\chi^2(4) = 5.56$, $p > .05$, and left the CFI (.96) unaltered, suggesting model equivalence. The chi-square test was again significant, $\chi^2(25) = 66.04$, $p < .01$; but the RMSEA indicated adequate fit (.064, 90% CI = .045–.083).

Discussion

Overall, our results suggest that interviews and checklists do not differ markedly in their quantification of youth externalizing symptoms in a general community sample, a pattern consistent with prior research (Gould et al., 1993; Jensen et al., 1996). In the community sample, constraining the associations between interview- and checklist-reported symptoms and teacher-reported externalizing symptoms to be equal did not reduce model fit, nor did constraining the associations between interview- and checklist-reported symptoms and the external validator variables. Notably, the internal-consistency reliability of each of the three scales was similar for the checklist and the interview – a likely explanation for the similarity of their associations to external variables.

In the clinical sample, the internal-consistency reliability of conduct disorder was markedly lower for the checklist than the interview. Consistent with this observation, the interview shared a stronger association with teacher-reported externalizing symptoms, and constraining this association to equal that shared by teacher-reported symptoms and the checklist yielded significantly worse model fit. It is important to note that the constrained model still provided adequate fit to the data; furthermore, when these constraints were imposed on the model assessing the associations between mother-reported symptoms and the three validator variables, the CFIs of the two models suggested equivalence.

In the clinical sample, then, the interview appeared to be 'performing' a bit better than the checklist, a difference that may have arisen because of enhanced reliability attributable to scale length (i.e., more items). To test this hypothesis, we reduced the DISC-IV scales to 'matched' items on the BCFPI and re-ran the models. In this secondary analysis, there was not a significant difference in fit between the unconstrained and constrained models, and neither the interview nor the checklist showed consistently stronger associations with the predictor variables. These findings suggest that when checklist scales are defined by only a few items, small performance differences between checklists and interviews may be due to scale length.

There were several limitations of the current study. First, in the community sample, the teachers used the same checklist as the mothers to report externalizing symptoms, which may have inflated the association between the two constructs. We attempted to offset this limitation by using two approaches to assessing equivalence, with the second approach dependent on the use of external 'validators' derived from objective indicators. Second, the pool of external validators available for study was limited for two reasons: (1) they could not be based on mothers' subjective reports of their own or their children's behavior; and (2) the selected variables needed to have established associations with externalizing symptoms. In addition, it was important to examine the same variables in both the community and clinic samples, to facilitate comparison between the models. The severity of the problems experienced by the clinical sample, as well as referral processes, may attenuate associations with the external variables. These factors may have contributed to the null associations between externalizing symptoms and both family income and gender in the clinic-referred

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

sample. The relatively weak associations with gender may also have been due, in part, to the inclusion of ODD symptoms, as data concerning gender differences in oppositional symptoms are equivocal (see Loeber, Burke, Lahey, Wingers, & Zera, 2000). It will be important in future to examine associations with other measures, such as biological indicators.

In summary, this study extended previous work by examining the equivalence of interview and checklist measures of youth psychopathology in both general community and clinically referred samples. Consistent with previous studies, we found no clear advantage for either approach in a community sample, indicating that it is not necessary to use the more onerous interview procedures to quantify psychopathology in more 'normative' samples. In the clinic sample, the interview performed marginally better than the very brief checklist; however, this advantage appeared to result from the greater number of items on the interview. Adding a few items to increase the reliable variance of very short checklists may strengthen such measures to the point of equivalence with interviews when quantifying psychopathology in clinic-referred samples. Only one previous study has addressed the issue of equivalence of interviews and checklists in a clinical sample (Ferdinand et al., 2004). In their analyses, the authors compared the predictive value of diagnostic classifications (present/absent) obtained with an interview to scale scores obtained with a checklist. Depending on the form of the associations with the dependent variables, the increased variability of the scale scores may have enhanced the predictive utility of this measure. Given the limited amount of work on this issue, it may still be premature to draw firm conclusions about comparability of interviews and checklists for the measurement of psychopathology in clinically referred youth until additional studies have compared interviews with longer rating scales. Future work should also examine the equivalence of checklists and interviewer-based interviews, as well as the equivalence of interviews and checklists for the quantification of internalizing symptoms.

Acknowledgements

Melanie Dirks was supported by an Ontario Mental Health Foundation post-doctoral fellowship during the preparation of this article. Michael Boyle holds a Canada Research Chair in the Social Determinants of Child Health

Correspondence to

Melanie A. Dirks, Department of Psychology, McGill University, Stewart Biology Building, Office W7/3J, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1, Canada; Tel: 514-398-3856; Email: melanie.dirks@mcgill.ca

Key points

- The few studies examining the equivalence of structured interviews and behavior checklists have found these instruments to provide comparable quantification of youth psychopathology.
- The current study found no marked performance difference between structured interviews and checklists in a community sample.
- In a clinic sample, the interview performed marginally better than the checklist, a difference that appeared to be due, in part, to the greater number of items on this measure.
- When quantifying youth externalizing symptoms in community samples, checklists and respondent-based interviews may provide researchers and clinicians with commensurate measurement.
- Further work should be done to assess the comparability of interviews and longer rating scales in clinical samples.

References

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

Angold, A., & Fisher, P.W. (1999). Interviewer-based interviews. In D. Shaffer, C.P. Lucas, & J.E. Richters, (Eds), *Diagnostic assessment in child and adolescent psychopathology* (pp. 34–63). New York: Guilford.

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

Boyle, M.H., Offord, D.R., Hofmann, H.F., Catlin, G.P., Byles, J.A., Cadman, D.T., Crawford, J.W., Links, P.S., Rae-Grant, N.I., & Szatmari, P. (1987). Ontario Child Health Study: I. Methodology. *Archives of General Psychiatry*, *44*, 826–831.

Boyle, M.H., Offord, D.R., Racine, Y., Fleming, J.E., Szatmari, P., & Sanford, M. (1993a). Evaluation of the revised Ontario Child Health Study scales. *Journal of Child Psychology and Psychiatry*, *34*, 189–213.

Boyle, M.H., Offord, D.R., Racine, Y., Sanford, M., Szatmari, P., Fleming, J.E., & Price-Munn, N. (1993b). Evaluation of the Diagnostic Interview for Children and

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

Adolescents for use in general population samples. *Journal of Abnormal Child Psychology*, *21*, 663–681.

Boyle, M.H., & Pickles, A.R. (1998). Strategies to manipulate reliability: Impact on statistical associations. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1077–1084.

Bradshaw, C.P., Buckley, J.A., & Ialongo, N.S. (2008). School-based service utilization among urban children with early onset educational and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly*, *23*, 169–186.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen, & J.S. Long, (Eds), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Collins, L.M., Schafer, J.L., & Kam, C.M. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods*, *6*, 330–351.

Costello, E.J., Angold, A., Burns, B.J., Erkanli, A., Stangl, D.K., & Tweed, D.L. (1996). The Great Smoky Mountains Study of Youth: Functional impairment and serious emotional disturbance. *Archives of General Psychiatry*, *53*, 1137–1143.

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

Duncan, G.J., Brooks-Gunn, J., & Klebanov, P.K. (1994). Economic deprivation and early-childhood development. *Child Development*, *65*, 296–318.

Ferdinand, R.F., Visser, J.H., Hoogerheide, K.N., van der Ende, J., Kasius, M.C., Koot, H.M., & Verhulst, F.C. (2004). Improving estimation of the prognosis of childhood psychopathology: Combination of DSM-III-R/DISC diagnoses and CBCL scores. *Journal of Child Psychology and Psychiatry*, *45*, 599–608.

Florsheim, P., Tolan, P.H., & Gorman-Smith, D. (1998). Single-parenthood, family factors and risk for behavior problems among African-American and Latino boys. *Child Development*, *5*, 1437–1447.

Gould, M.S., Bird, H., & Staghezza, J.B. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. *Journal of Abnormal Child Psychology*, *21*, 287–313.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jacobusse, G. (2005). *WinMice user's manual for prototype version 0.1*. Netherlands: TNO.

Jensen, P.S., Watanabe, H.K., Richters, J.E., Roper, M., Hibbs, E.D., Salzberg, A.D., & Liu, S. (1996). Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *Journal of Abnormal Child Psychology*, *24*, 151–168.

Kline, R.B. (2005). *Principles and practices of structural equation modeling*. New York: Guilford.

Loeber, R., Burke, J.D., Lahey, B.B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: A review of the past 10 years, Part I. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 1468–1484.

MacCallum, R.C., Browne, M.W., & Sugara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.

Reich, W., & Welner, Z. (1988). *Revised version of the Diagnostic Interview for Children and Adolescents (DICA-R)*. St. Louis, MO: Department of Psychiatry, Washington University School of Medicine.

Rey, J.M., Walter, G., & Soutullo, C.A. (2007). Oppositional defiant and conduct disorders. In A. Martin, & F.R. Volkmar, (Eds.), *Lewis' child and adolescent psychiatry: A comprehensive textbook* (pp. 454–466). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Roberts, R.E., Attkisson, C.C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry*, *155*, 715–725.

Shaffer, D., Fisher, P., & Lucas, C.P. (1999). Respondent-based interviews. In D. Shaffer, C.P. Lucas, & J.E. Richters, (Eds), *Diagnostic assessment in child and adolescent psychopathology* (pp. 3–33). New York: Guilford.

Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., & Schwab-Stone, M.E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38.

Spetie, L., & Arnold, E.L. (2007). Attention-deficit/hyperactivity disorder. In A. Martin, & F.R. Volkmar, (Eds), *Lewis' child and adolescent psychiatry: A comprehensive textbook* (pp. 430–453). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69.

Manuscript accepted February 2010

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

© 2010 The Authors
Journal compilation © 2010 Association for Child and Adolescent Mental Health.

Appendix: Zero-order correlations among all measured variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. ADHD –Interview	–	.74**	.44**	.35**	.51**	.36**	.35**	.09	.15**	−.20**	−.34**	−.08	−.03	.89**	.39**	.46**
2. ADHD – Checklist	.71**	–	.38**	.41**	.45**	.44**	.30**	.11*	.14**	−.14**	−.30**	−.04	−.12*	.71**	.32**	.42**
3. CD – Interview	.46**	.45**	–	.68**	.58**	.47**	.22**	.35**	.23**	−.10**	−.20**	−.08	−.03	.43**	.83**	.47**
4. CD – Checklist	.33**	.59**	.54**	–	.48**	.47**	.19**	.22**	.14**	−.08	−.17**	−.04	−.16*	.34**	.67**	.37**
5. ODD –Interview	.62**	.56**	.63**	.58**	–	.67**	.23**	.25**	.21**	−.04	−.19**	−.01	−.10	.43**	.54**	.89**
6. ODD – Checklist	.48**	.72**	.57**	.71**	.73**	–	.13**	.16**	.17**	.06	−.08	−.06	−.12*	.30**	.39**	.65**
7. ADHD – Teacher	−.39**	−.40**	−.33**	−.27**	−.27**	−.26**	–	.55**	.69**	−.28**	−.47**	.04	.01	.26**	.14**	.19**
8. CD – Teacher	−.32**	−.27**	−.33**	−.22**	−.31**	−.24**	.52**	–	.74**	−.21**	−.25**	−.02	−.04	.04	.25**	.17**
9. ODD – Teacher	−.36**	−.37**	−.39**	−.36**	−.38**	−.35**	.59**	.74**	–	−.23**	−.22**	−.05	.05	.07	.11*	.16**
10. Gender	−.12	−.19**	−.22**	−.20**	−.01	−.14*	.23**	.17**	.25**	–	.15**	.01	−.03	−.17**	−.10*	.02
11.Academic performance	−.30**	−.21**	−.24**	−.23**	−.29**	−.25**	.49**	.22**	.27**	.18**	–	.05	−.07	−.36**	−.14**	−.13*
12. Family income	−.21**	−.23**	−.21**	−.28**	−.17**	−.23**	.20**	.04	.13*	.09	.22**	–	.24**	−.05	−.08	−.02
13. Single parent	−.21**	−.13*	−.20**	−.11	−.12	−.17**	.13*	.10	.05	.13*	.26**	.58**	–	−.01	−.09	−.13*
14. ADHD – interview reduced scale	–	–	–	–	–	–	–	–	–	–	–	–	–	–	.38**	.39**
15. CD – interview reduced scale	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	.43**
16. ODD – interview reduced scale	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Note: Data from the community sample is presented below the diagonal, data from the clinic sample is presented above the diagonal. Reduced interview scales were not calculated for the community sample, as the checklist and the interview consisted of comparable numbers of items. In the community sample, an inverse transformation was applied to the distributions for teacher-reported ADHD, CD, and ODD; thus, higher scores reflect fewer symptoms. Gender dummy coded males = 0. Single-parent status dummy coded single-parent family = 0. ADHD = attention deficit hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder.
*$p < .05$; **$p < .01$.