The 2014 Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*) Part II: Psychometric Adequacy for Categorical Measurement of Selected DSM-5 Disorders

Michael H. Boyle, PhD, Professor Emeritus, McMaster University[1,2]
Laura Duncan, MA, PhD Student and Research Coordinator, McMaster University[1,2]
Katholiki Georgiades, PhD, Associate Professor, McMaster University[1,2]
Li Wang, MSc, PhD Student and Analyst, McMaster University[1,2]
Jinette Comeau, PhD, Assistant Professor, King's University College at Western University[5]
Mark A. Ferro, PhD, Assistant Professor, University of Waterloo[6]
Ryan J. Van Lieshout, MD, PhD, FRCPC, Assistant Professor, McMaster University[1,2]
Peter Szatmari, MD, MSc, FRCPS, Professor and Head, Centre for Addiction and Mental Health, Hospital for Sick Children, University of Toronto [1,7]
Harriet L. MacMillan, MD, MSc, FRCPC, Professor, McMaster University[1,2,3]
Kathryn Bennett, PhD, Professor, McMaster University[1,4]
Magdalena Janus, PhD, Associate Professor, McMaster University[1,2]
Ellen L. Lipman, BSc, MD, FRCP(C), MSc, Professor and Director, McMaster University [1,2]
Anna Kata, MA, MPH, Research Coordinator, McMaster University[1,2]
[1]Offord Centre for Child Studies; [2]Department of Psychiatry and Behavioural Neurosciences; [3]Department of Paediatrics; [4]Department of Health Research Methods, Evidence and Impact (formerly Clinical Epidemiology and Biostatistics); McMaster University, 1280 Main St W, Hamilton, ON, L8S 4K1, Canada
[5]Department of Sociology, King's University College at Western University, 266 Epworth Ave, London, ON, N6A 2M3, Canada

[6]School of Public Health and Health Systems, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
[7]Department of Psychiatry, University of Toronto, 27 King's College Circle, Toronto, ON, M5S 1A1, Canada

Correspondence concerning this article should be addressed to: Laura Duncan, Offord Centre for Child Studies, McMaster University, 1280 Main St, West, MIP 201A, Hamilton Ontario, L8S 4K1. Email: duncanlj@mcmaster.ca Phone: 1 905 525 9140 x21504 Fax: 905 574 6665

Abstract

Objectives: To compare the reliability and convergent validity of parent assessments from the Mini International Neuropsychiatric Interview for Children and Adolescents (*MINI-KID*—a structured diagnostic interview) and the Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*) symptom checklist for classifying conduct disorder (CD), conduct disorder or oppositional defiant disorder (CD-ODD), attention deficit hyperactivity disorder (ADHD), major depressive disorder (MDD), generalized anxiety disorder (GAD) and separation anxiety disorder (SAD) based on DSM-5 criteria.

Methods: Data came from 283 parent–youth dyads aged 9 to 18 years. Parents and youth completed the assessments separately on two different occasions 7 to 14 days apart. After converting the *OCHS-EBS* scale scores to binary disorder classifications, we compare test-retest reliability estimates and use structural equation modelling (SEM) to compare estimates of convergent validity for the same disorders assessed by each instrument.

Results: Average test-retest reliabilities based on kappa were 0.71 (*MINI-KID*) and 0.67 (*OCHS-EBS*). The average β coefficients for three latent measures comprised of the following indicators–(1) parent perceptions of youth mental health need and impairment; (2) diagnosis of specific disorders based on health professional communications and youth taking prescribed medication; and (3) youth classifications of disorder based on the *MINI-KID*–were 0.67 (*MINI-KID*) and 0.69 (*OCHS-EBS*).

Conclusion: The *OCHS-EBS* and *MINI-KID* achieve comparable levels of reliability and convergent validity for classifying child psychiatric disorder. The flexibility, low cost and minimal respondent burden of checklists for classifying disorder make them well suited for studying disorder in the general population and screening in clinical settings.

Keywords Symptom checklist, Structured diagnostic interview, Measurement, Structural equation modelling, Validity, Reliability, Child psychiatric disorder

Introduction

Reliable, valid and inexpensive instruments are needed to measure child and adolescent (youth) psychiatric disorder conceptualized as both dimensional and categorical (present or absent) phenomena for use in epidemiological studies in the general population and screening in clinical settings.[1] The most common approaches used to measure youth disorders are structured and semi-structured standardized diagnostic interviews (SDIs) and self-completed symptom checklists.[2,3] Interviews focus on disorder as a categorical phenomenon, drawing on symptom and impairment criteria specified in the Diagnostic and Statistical Manual of Mental Disorders (DSM)[4] to classify disorder. Checklists focus on mental problems as dimensional phenomena, drawing on empirical methods such as factor analysis to identify syndromes based on parent or youth ratings of problem behaviours.

SDIs are expensive and time consuming to implement. For example, the Diagnostic Interview Schedule for Children 4th Edition takes on average 70 minutes to complete for a non-clinic respondent (general population) and 90-120 minutes for a clinic respondent.[5] To lessen the burden of response, most interviews use screening questions to skip respondents out of modules where they are likely to test negative.[6] This strategy leads to information loss about psychiatric symptoms and an ability to construct dimensional measures of disorder applicable to all respondents. Checklists are brief, simple, inexpensive to implement, pose little burden to respondents and collect information on all symptoms. Choosing cut-points along the continuum of scale scores allows checklists to represent disorder categorically as well as dimensionally. Demonstrating comparable reliability and validity between checklists and SDIs would greatly expand our ability to study and screen for child psychiatric disorder in situations where SDIs would be too burdensome (general population studies, community child mental health centres).

Although SDIs have become the de facto gold standard for classifying youth psychiatric disorder,[7] there are compelling arguments for expecting checklists to classify psychiatric disorder as reliably and validly as SDIs.[8] Admittedly, the empirical studies[9-14] are dated, few in number and associated with some important limitations that include: (1) relatively small samples; (2) lack of comparative data on the test-retest reliabilities of the instruments; (3) inattention to prevalence effects; (4) reliance on subjective interpretation of the numerical findings in the absence of formal empirical tests; and (5) failure to account for measurement error in comparing the validity of the instruments.

In this study, we conduct a direct comparison of the reliability and validity of the Ontario Child Health Study Emotional Behavioural Scales (*OCHS-EBS*)[15] measuring conduct disorder (CD), conduct disorder or oppositional-defiant disorder (CD-ODD), attention-deficit hyperactivity disorder (ADHD), generalized anxiety disorder (GAD), separation anxiety disorder (SAD) and major depressive disorder (MDD) with the parent Mini International Neuropsychiatric Interview for Children and Adolescents (*MINI-KID-P*—a structured diagnostic interview). (CD and ODD are combined because the *MINI-KID* skips respondents over ODD when youth test positive for CD.)[16,17] We address the limitations and extend previous studies by: (1) comparing the test-retest reliability of the two instruments for classifying disorder in the same time interval (7-14 days); (2) implementing formal empirical tests of convergent validity; (3) using structural equation modelling (SEM) with latent variables free of measurement error for the validity analysis; and (4) conducting a sensitivity analysis to determine the extent to which instrument differences in prevalence account for differences in convergent validity.

Methods

Participants

Two hundred and eighty-three parent–youth dyads aged 9 to 18 years (185 from the general population and 98 from a mental-health outpatient clinic) participated. One parent participant, in addition to 5 parent-youth dyad participants (2.1%), did not complete the retest interview and were removed from the reliability analysis.

GENERAL POPULATION PARTICIPANTS

Youth in the general population were sampled from 4 elementary schools (grades 5-8) and 4 high schools (grades 9-12), enlisted by school board representatives. Students took home a study letter, a consent form to be completed by parents and a 7-item screening questionnaire to be completed by parents of elementary students or students themselves, if attending high school (N=4,333). Students who returned signed parental consent agreeing to be contacted about the study and completing the screening questionnaire (n=1,210) formed the eligible sample (27.9% response). The questionnaire had seven items that included assessments of student's emotional, social and academic functioning. The items, identical for parents and youth, were scored from positive to negative and summed to produce a distribution of risk. Based on parent assessments (elementary school) or youth assessments (secondary school), youth were classified at high risk (top 10%), medium risk (11-30%) or low risk (bottom 70%), sampled in equal numbers from each risk group and invited to participate. Youth classified at high and medium risk were oversampled to increase the number of participants likely to have disorder. Across the strata, 346 were sampled and 185 participated—34.1% at low risk, 30.8% at medium risk, and 35.1% at high risk. Sampling weights were created based on the probability of youth being selected and participating within each stratum.

MENTAL HEALTH OUTPATIENT CLINIC PARTICIPANTS

Eligible were families who provided informed consent; had youth aged 10 to 17 at no immediate risk of self-harm or harm to others and exhibited no apparent developmental or learning problem such as autism or a learning disability. One university and one community-based children's mental health centre contributed data to the clinic sample. In the university-based centre, 243 families seen at intake during the study period were deemed eligible. The research team contacted 129 of these families (53.1%) and 54 participated (22.2%). In the community-based centre, 158 families seen at intake during the study period were deemed eligible. The research team contacted all of these families and 45 participated (28.5%).

Families were interviewed 7-14 days apart between December 2011 and December 2013. All study procedures, including consent and confidentiality requirements, were approved by the Hamilton Integrated Research Ethics Board at McMaster University, the Research Ethics Committees at the School Boards and the clinics involved in the study.

Concepts and Measures

MEASURES OF PSYCHIATRIC DISORDER

MINI-KID-P. The *MINI-KID-P* for parents and *MINI-KID-Y* for youth are SDIs that assess DSM-IV-TR disorders in youth aged 6 to 17 years.[17] Validated against the Schedule for Affective Disorders and Schizophrenia for School Aged Children-Present and Lifetime Version, the *MINI-KID* has a 1-to 5-day test-retest reliability based on kappa (>0.75) for all diagnoses identified in combined interviews with parents and youth.[17] In our study, the interviews were administered separately: the *MINI-KID-Y* to youth; and the *MINI-KID-P* to parents.[15]

OCHS-EBS. The scales used in this study were developed for the 2014 Ontario Child Health

Study to provide both dimensional and categorical representations of child psychiatric disorder.[15] They draw on items used in our previous studies[18,19] as well as new items judged by clinicians and researchers to approximate DSM-5 criteria. The reference period for assessing items is the past 6 months, and each one is scored 0, 1, or 2, indicating responses of "never or not true," "sometimes or somewhat true," and "often or very true," respectively. The raw scores are summed to form a scale score to measure each disorder. The *OCHS-EBS* takes about 8-10 minutes for a parent to complete.

CONVERGENT VALIDITY INDICATORS

Convergent validity is an approach to testing the validity of a measure by quantifying its strength of association with measures of similar constructs hypothesized to be linked theoretically. In this study, we compare the strength of association between the classifications of disorder based on the *MINI-KID-P* and *OCHS-EBS* with three highly related constructs measured as latent variables. Variable one is parental perceptions of their youth's need for professional help with emotional or behavioural problems, in conjunction with impaired social or academic functioning—a general trait assessed in the last 6 months and hypothesized to underlie all types of psychiatric disorder.[20] Variable two includes specific types of emotional-behavioural problems ever diagnosed by health care professionals or school personnel and communicated to parents, in conjunction with parental reports of their youth currently taking prescribed medication for the same problem. Variable three is youth classifications of the same disorders identified independently by the *MINI-KID-Y*.

Mental Health Need/Impairment: *Parent Ratings.*
Need for Help: A binary indicator coded positive (1) when respondents checked yes to two questions, "During the last 6 months, do you think that ___ has had any emotional or behavioural problems? Do you think that ___ needs or needed professional help with these problems?"
Impaired Social Functioning: A summated rating scale of items scored from (1) very well, no problems, to (5) not well at all, constant problems, in response to the stem question: "During the past 6 months, how well has ___ gotten along with..." asking about (a) other kids such as friends or classmates; (b) teachers at school; and (c) the family.
Impaired Academic Functioning: A single item scored from (1) excellent student to (5) poor student, constant problems in response to the question, "Which of the statements best describes how well ___ has done overall in subjects at school during the past 6 months?"

Health Care Diagnosed Problems and Use of Prescription Medication: *Parent Report.*
Specific Diagnoses of Emotional-behavioural Problems Lifetime: Binary classifications derived from positive responses to the following question, "Have you ever been told by a teacher, school official, doctor, nurse or other health professional that ___ has: a) anxiety; b) depression; c) attention problems; d) behavioural problems?"
Use of Prescription Medications Currently: Binary classifications derived from positive responses to a stem question and follow-up questions: "Is ___ currently taking any prescribed medication?" and "What does ___ take this medication for: Hyperactivity? Behavioural problem? Depression? Anxiety?"

Youth (Cross-informant) Classifications of the same Disorders based on the MINI-KID-Y.
Please see *MINI-KID-P* above.
Analyses

CONVERTING OCHS-EBS SCALE SCORES TO BINARY CLASSIFICATIONS OF DISORDER

Scale scores on the *OCHS-EBS* were converted to binary measures of disorder independently of the *MINI-KID-P* at the thresholds matching the general population prevalence estimates for CD (2.1%); ODD (3.6%); ADHD (3.4%) and MDD (1.3%) identified among youth in a recent worldwide meta-analysis of prevalence studies.[21] Prevalence estimates for GAD (1.8%) and SAD (1.9%) were taken from a different review.[22] Each scale score was converted to a binary measure at the threshold closest to the prevalence of its corresponding disorder identified above. These threshold scores were determined in the weighted general population sample (see General Population Participants) and applied to all respondents.

PREVALENCE, TEST-RETEST RELIABILITY AND CROSS-INSTRUMENT AGREEMENT

The six-month prevalence of disorders assessed by the *MINI-KID-P* and *OCHS-EBS* is expressed as a percent. Test-retest reliability over a 1-to 2-week period and cross-instrument agreement are estimated by kappa.[23] Our sample size for reliability is 277 parent–youth dyads because 6 did not complete the retest. With a type I error (alpha) set at 0.05 (2-tailed), the statistical power (1-beta) available in the study to identify a difference in kappa between the *MINI-KID-P* and *OCHS-EBS* of |0.20| goes from about 35% to 95%. This variability in power depends on the test-positive rate (prevalence) expected to go from 0.04 to 0.26.

CONVERGENT VALIDITY

We used SEM to test for differences in convergent validity between the *MINI-KID-P* and *OCHS-EBS* classifications of disorder. SEM is a multivariate statistical technique with a measurement component—derivation of latent variable measures based on indicator variables—and a structural component—the specification of relationships among the latent variable measures. To remove temporal error, we construct latent variable measures of each disorder (dependent variables) for each instrument based on their assessments at each time point. We also construct latent variable measures of our convergent validity variables (independent variables). For 2 of the variables—youth mental health need/impairment and health care diagnosed problems and use of prescription medication—the latent variable measures are based on their time 1 indicators because these questions were not repeated at time 2. For classifications of the same disorders based on the *MINI-KID-Y*, we create latent variable measures of disorder based on their assessments at each time point as we did with the *MINI-KID-P*.

To compare the convergent validity of the *MINI-KID-P* and *OCHS-EBS* classifications of disorder, we specify separate regression models for each disorder and test for differences in the magnitude (*β* coefficients) and strength (explained variance) of association linking the *MINI-KID-P* and *OCHS-EBS* to each of the construct validity variables. Each model consists of three latent variable measures: one each for the two instruments (dependent variables), and one for the convergent validity variable (independent variable). Figure 1 illustrates the regression of the *MINI-KID-P* and *OCHS-EBS* classifications of ADHD on *MINI-KID-Y* classifications of ADHD.

We used MPlus 7.4[24] to develop separate SEMs for each disorder. MPlus offers a generalized measurement component which allows for dichotomous and ordered categorical variables (indicators) in the derivation of latent variable measures.[25] Adequate model fit was defined as values ≥0.98 for the comparative fit index (CFI, range 0 to 1.0); ≤0.05 for the root-mean-squared error of approximation (RMSEA); and a non-significant chi-square for model fit. The Wald statistic follows the Student's *t* distribution.[26] Using maximum likelihood to handle missing retest information, our sample size for validity is 283. With type I error (alpha) set at

0.05 (2-tailed), power depends on the effect size (difference between $\beta_{MINI\text{-}KID\text{-}P}$ and $\beta_{OCHS\text{-}EBS}$) and the standard deviation of the regression errors. In our study, these errors varied at the extremes between 0.38 and 0.95 (standardized) so that effect size differences that go from about 0.07 to 0.24 can be reliably identified with 80% power.

Finally, we conducted a sensitivity analysis to determine if statistically significant differences in convergent validity between the *MINI-KID-P* and *OCHS-EBS* were influenced by our approach to setting thresholds for classification. To do this, we re-ran the original SEM analysis after re-setting the *OCHS-EBS* thresholds for classifying disorder to align with the prevalence estimates observed for the *MINI-KID-P* in our general population sample.

Results

The sample characteristics and distribution of the construct validity indicators appear in Table 1. There are fewer males (43.5%) than females. The average age of youth is 14.8 (SD=2.3) years.

In Table 2, the weighted prevalence of disorders identified by the *OCHS-EBS* approximate the population estimates on which they are based and are similar to the *MINI-KID-P* except for MDD where the prevalence is 6.9 versus 2.0% for the *MINI-KID-P* and *OCHS-EBS*, respectively. The unweighted prevalence estimates for the general population and clinic samples combined are higher for CD and ADHD based on the *OCHS-EBS* and for CD-ODD, GAD and MDD based on the *MINI-KID-P*.

Differences between instruments in the test-retest reliability based on kappa are ≤|0.10| for all of the disorders except for GAD (0.19 higher in the *MINI-KID-P*) and SAD (0.12 higher in the *OCHS-EBS*) – and none of the differences are statistically significant. Based on kappa, agreement between instruments on the classifications of disorder goes from 0.38 (MDD, SAD) to 0.59 (CD-ODD).

Table 3 shows the SEM results used to test equivalence in the convergent validity of the instruments based on the construct validity variables. Each line represents a separate SEM and all models provide excellent fit based on the CFI (all ≥0.98) and RMSEA (all ≤0.05) (not shown). All estimates based on the model fit χ^2 are non-significant at $P>0.05$.

Among the 18 SEMs in Table 3, the β coefficients are numerically larger for the *MINI-KID-P* in 7 comparisons, for the *OCHS-EBS* in 10 comparisons, and identical in one comparison. The average size of the β coefficients for the interview and checklist are: 0.67 and 0.69, respectively. The Wald tests of parameter constraints with one degree of freedom indicate that constraining the unstandardized b coefficients or residual variances to be equal led to statistically significant loss of fit ($\chi^2 ≥3.86$, $P<0.05$) in 5 comparisons and marginally significant loss of fit ($P≥0.05$, <0.10) in 4 comparisons. Although there is no obvious pattern of between-instrument differences, it appears that the *MINI-KID-P* is stronger in measuring GAD while the *OCHS-EBS* may be stronger in measuring CD.

Table 4 shows the effect of re-setting the checklist thresholds to align with the prevalence of disorder observed for the *MINI-KID-P*. This analysis is restricted to the 8 disorders associated with significant loss of fit. In comparison with the $\beta_{OCHS\text{-}EBS}$ in Table 3, the $\beta_{OCHS\text{-}EBS}$ in Table 4 converges towards $\beta_{MINI\text{-}KID\text{-}P}$ in all instances. With two exceptions—CD-ODD (diagnostic groupings) and CD (youth identified psychiatric disorder)—all of the effects in Table 3 were rendered statistically non-significant.

Discussion

This study indicates that a self-administered problem checklist can achieve the same levels of reliability and convergent validity for classifying youth psychiatric disorder as a structured SDI. These findings are consistent with the small number of investigations done in the 1990s that examined the construct validity of interviews and checklists for classifying child psychiatric disorder.[9-14]

There are important challenges associated with comparing the psychometric properties of SDIs and checklists. The first challenge arises from limits to our understanding about the nature of child psychopathology. In the absence of criterion measures, we rely on construct validity indicators to assess the comparative validity and usefulness of these instruments. These indicators should be theoretically important, empirically supported (reliable and valid), and 'independent' of the specific questions/items and methods making up the competing instruments. Although our third approach to convergent validity—use of using cross-informant classifications of disorder based solely on the *MINI-KID-Y*—violates one of these dictums, we believe that the absence of between-instrument differences in their strength of association with *MINI-KID-Y* classifications of disorder is strong evidence of their equivalence.

The second challenge arises from selecting checklist thresholds for identifying disorders. To assure independence of the *MINI-KID-P*, we aligned our checklist thresholds with population prevalence estimates from a meta-analysis. The few convergent validity advantages of either instrument can be traced to differences in prevalence which, in turn, can be traced to differences in reliability, particularly at the extremes of prevalence.[27] A recent study of three SDIs yielded prevalence estimates of 1+ disorders in the same respondents of 47.1%, 32.4% and 17.7%[28] illustrating that the challenge of selecting thresholds is not unique to checklists.

The third challenge focuses on a trio of methodological concerns arising from sampling and response, statistical power and measurement error. One, comparative studies should be done separately in clinical and general population samples—a requirement far beyond our funding capacity. Our sampling strategy reflected a desire to represent the general population while ensuring that there would be enough youth classified with disorder to conduct meaningful convergent validity analyses. Without access to information on non-respondents, we cannot evaluate the representativeness of our samples. If there are selection factors at work in our study, they would need to exert a differential effect on reliability and validity across instruments which seems unlikely to us. Two, large samples are needed to have adequate statistical power for comparing the psychometric properties of different instruments measuring the same traits.[29] Although our sample is large compared with other studies, it is still limited. Furthermore, power in a given study will vary across disorders because of differences in their prevalence. Three, uncontrolled measurement error is a serious threat when comparing the validity of measurement instruments. The use of SEM to remove measurement error in the classification of disorder and measurement of convergent validity variables substantially enhanced our ability to conduct meaningful tests. However, the use of SEM did not fully overcome the effects of prevalence differences on associations between the convergent validity variables and disorders. Not taking prevalence differences into account could lead one to believe mistakenly that the validity and usefulness of alternative instruments for classification might depend on the type of disorder being assessed.

Checklists versus Interviews

In this paper, we focus exclusively on the measurement objective of classifying disorder for epidemiological studies in the general population and screening in clinical settings in a head-

to-head comparison with structured SDI. In clinical settings, semi-structured SDIs serve the broader diagnostic objectives of engaging patients and formulating an intervention plan. This process depends on years of clinical training and experience. Although checklists can contribute to this process through screening, they cannot substitute for it.

Over the past 30 years, substantial resources have gone into the development of structured SDIs, resulting in a belief of their superiority. At the same time, there is a willingness to overlook differences among them in prevalence arising from the same diagnostic criteria[28] and the fact that the overall test-retest reliability of SDIs is modest at best (κ=0.58 (CI 95% 0.53-0.63)) and highly variable across studies.[30] Given the striking differences in cost and burden between structured SDIs and checklists, it is surprising how little research has been directed towards examining their relative scientific merits. In our view, carefully developed symptom checklists can substitute for structured SDIs and provide an effective way to measure child and youth psychiatric disorder as both categorical and dimensional phenomena. Studies addressing this question are urgently needed to provide researchers and clinicians with an appropriate evidence base for making cost-effective decisions about using checklists or SDIs for classifying youth disorder in epidemiological studies and for screening in clinical practice.

References

1. Coghill D, Sonuga-Barke EJS. Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders – implications of recent empirical study. J Child Psychol Psychiatry. 2012;53(5):469–489.

2. Angold A. Chapter 3. Diagnostic Interviews with Parents and Children. In: Rutter M, Taylor E, editors. Child and adolescent psychiatry. 4th ed. Oxford (UK): Blackwell Publishing Inc; 2002. p. 32–51.

3. Verhulst FC, Van der Ende J. Chapter 5. Rating scales. In: Rutter M,Taylor E, editors. Child and adolescent psychiatry. 4th ed. Oxford (UK): Blackwell Publishing Inc; 2002. p. 70–86.

4. American Psychiatric Association. Diagnostic and statistical manual of mental Disorders. 5th ed. Arlington (VA): American Psychiatric Publishing; 2013.

5. Shaffer D, Fisher P, Lucas CP, et al. National Institute of Mental Health Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. J Am Acad Child Adolesc Psychiatry. 2000;39(1):28–38.

6. Kessler RC, Wittchen H-U, Abelson JM et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. Int J Methods Psychiatr Res. 1998;7(1):33–55.

7. Rettew DC, Lynch AD, Achenbach TM, et al. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int J Methods Psychiatr Res. 2009;18(3):169–184.

8. Boyle MH, Duncan L, Georgiades K, et al. Classifying child and adolescent psychiatric disorder by problem checklists and standardized interviews. Int J Methods Psychiatr Res. 2016;26(4):e1544.

9. Boyle MH, Offord DR, Racine Y, et al. Interviews versus checklists: adequacy for classifying childhood psychiatric disorder based on adolescent reports. Int J Methods Psychiatr Res. 1996;6:309–319.

10. Boyle MH, Offord DR, Racine Y, et al. Adequacy of interviews versus checklists for classifying childhood psychiatric disorder based on parent reports. Arch Gen Psychiatry. 1997;54(9):793–797.

11. Dirks MA, Boyle MH. The comparability of mother-report structured interviews and checklists for the quantification of youth externalizing symptoms. J Child Psychol Psychiatry. 2010;51(9):1040–1049.

12. Gould MS, Bird H, Jaramillo S. Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. J Abnorm Child Psychol. 1993;21(3):287–313.

13. Jensen PS, Watanabe HK, Richters JE, et al. Scales, diagnosis and child psychopathology: II. Comparing the CBCL and the DISC against external validators. J Abnorm Child Psychol. 1996;24(2):151–168.

14. Jensen PS, Watanabe HK. Sherlock Holmes and child psychopathology assessment approaches: The case of the false-positive. J Am Acad Child Adolesc Psychiatry. 1999;38(2):138–146.

15. Duncan L, Georgiades K, Wang L, et al. The 2014 Ontario Child Health Study Emotional Behavioural Scales (OCHS-EBS) Part I: a checklist for dimensional measurement of selected DSM-5 disorders. Submitted.

16. Duncan L, Georgiades K, Wang L, et al. Psychometric evaluation of the Mini International

Neuropsychiatric Interview for Children and Adolescents (MINI-KID). Psychol Assess. 2017;30(7):916–928.

17. Sheehan DV, Sheehan KH, Shytle RG, et al. Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). J Clin Psychiatry. 2010;71(3):313–326.

18. Boyle, M, Offord DR, Hofmann HF, et al. Ontario Child Health Study: I. methodology. Arch Gen Psychiatry. 1987;44(9):826–831.

19. Boyle MH, Offord DR, Racine YA, et al. Evaluation of the revised Ontario Child Health Study Scales. J Child Psychol Psychiatry. 1993;34(2):189–213.

20. Wittchen HU, Nelson CB, Lachner G. Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. Psychol Med. 1998;28:109–126.

21. Polanczyk GV, Salum GA, Sugaya LS, et al. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015;56(3):345–365.

22. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci. 2009;11:7–20.

23. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas, 1960;20:37–46.

24. Muthén LK, Muthén BO. Mplus User's Guide. 6th ed. Los Angeles (CA): Muthén & Muthén; 2016

25. Muthen B. A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. Psychometrika. 1984;4:115–132.

26. Dupont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. Control Clin Trials. 1998;19:589–601.

27. Shrout P. Measurement reliability and agreement in psychiatry. Stat Methods Med Res. 1998;7:301–317.

28. Angold A, Erkanl, A, Copeland W, et al. Psychiatric diagnostic interviews for children and adolescents: A comparative study. J Am Acad Child Adolesc Psychiatry. 2012;38:506–517.

29. Lin HM, Williamson JM, Lipsitz SR. Calculating power for the comparison of dependent κ-coefficients. J R Stat Soc Ser C Appl Stat. 2003;52:391–404.

30. Duncan L, Comeau J, Wang L, et al. Research Review: Test-retest reliability of standardized diagnostic interviews to assess child and adolescent psychiatric disorders: a systematic review and meta-analysis. J Child Psychol Psychiatry. 2018. https://doi.org/10.1111/jcpp.12876 [Epub ahead of print]

Table 1. Sample characteristics and convergent validity variables (*n*=283)

Sample Characteristics	%	Mean (SD)
Youth		
Male	43.5	
Mean age, years		14.8 (2.3)
Born > Canada	8.5	
Parent		
Birth mother	83.0	
Mean age, years		44.7 (6.8)
Born > Canada	19.1	
≤Secondary education	25.8	
Family		
Lone parent	29.3	
Mean income $'000s		72.9 (39.6)
Convergent Validity Indicators		
Mental Health Need/Impairment		
Need for Help	42.4	
Mean impaired academic functioning		2.37 (1.08)
Mean impaired social functioning		6.32 (2.40)
Diagnostic groupings		
Specific mental health problems		
a) Behaviour	25.8	
b) Attention	36.7	
c) Depression	21.2	
d) Anxiety	36.4	
Prescribed medication		
a) Behaviour	8.5	
b) Hyperactivity	12.4	
c) Depression	9.5	
d) Anxiety	13.8	
Youth Psychiatric Disorder *MINI-KID*		
CD	2.5	
CD-ODD	11.7	
ADHD	4.2	
MDD	15.9	
GAD	12.4	
SAD	3.2	

Note: ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; CD-ODD: conduct disorder or oppositional defiant disorder; GAD: generalized anxiety disorder; MDD: major depressive disorder; *MINI-KID:* Mini International Neuropsychiatric Interview for Children and Adolescents; SAD: separation anxiety disorder.

Table 2. Six-month prevalence and 1-2 week test-retest reliability of DSM-IV disorders based on parent assessments obtained by *MINI-KID-P* and *OCHS-EBS*

Disorder	Weighted Prevalence[a] (*n*=185)		Unweighted Prevalence (*n*=283)		Test-retest Reliability (*n*=277) κ (SE)		*MINI-KID-P OCHS-EBS* Agreement (SE) (*n*=283)
	MINI-KID-P	*OCHS-EBS*	*MINI-KID-P*	*OCHS-EBS*	*MINI-KID-P*	*OCHS-EBS*	
CD	0.7	1.7	6.4	13.1	.67 (.10)	.65 (.08)	.46 (.09)
CD-ODD	3.6	4.1	26.5	18.4	.77 (.04)	.73 (.06)	.59 (.06)
ADHD	2.7	2.4	9.5	18.4	.77 (.07)	.71 (.06)	.49 (.07)
GAD	3.1	2.1	19.8	9.5	.75 (.05)	.56 (.09)	.46 (.07)
MDD	6.9	2.0	20.1	7.1	.67 (.06)	.62 (.09)	.38 (.07)
SAD	1.3	2.7	4.6	12.4	.60 (.12)	.72 (.07)	.38 (.09)

[a] Based on general population sample responses weighted inversely to their probability of being selected. All other estimates based on combined, unweighted general and clinic population sample response.

Note: ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; CD-ODD: conduct disorder or oppositional defiant disorder; GAD: generalized anxiety disorder; MDD: major depressive disorder; *MINI-KID-P:* Mini International Neuropsychiatric Interview for Children and Adolescents Parent Version; *OCHS-EBS:* Ontario Child Health Study Emotional Behavioural Scales; SAD: separation anxiety disorder.

Table 3. SEM regressions of parent latent variable measures of disorder (*MINI-KID-P* and *OCHS-EBS*) on latent factor predictors

	β coefficient (se)		Model fit $\chi^2_{(df)}$	Wald χ^2 (*P-value*)	
	$\beta_{MINI\text{-}KID\text{-}P}$	$\beta_{OCHS\text{-}EBS}$	*P-value*	*b*	Residual
Covariate					
Disorder					
Mental Health					
Need/Impairment					
CD	.73 (.08)	.85 (.05)	16.03 (11) *.14*	3.69 (.06)*	1.66 (.20)
CD-ODD	.93 (.03)	.92 (.04)	12.99 (11) *.29*	0.16 (.69)	0.15 (.70)
ADHD	.76 (.07)	.90 (.05)	12.39 (11) *.34*	1.77 (.18)	4.32 (.04)†
MDD	.82 (.05)	.79 (.09)	17.68 (11) *.09*	0.03 (.86)	0.18 (.67)
GAD	.79 (.05)	.65 (.08)	18.87 (11) *.06*	5.84 (.02)†	1.01 (.32)
SAD	.60 (.10)	.55 (.08)	15.81 (11) *.15*	0.42 (.52)	0.01 (.93)
Diagnostic Groupings					
CD	.63 (.11)	.65 (.10)	4.11 (6) *.66*	0.07 (.80)	0.00 (.95)
CD-ODD	.81 (.06)	.70 (.07)	4.62 (6) *.59*	1.57 (.21)	2.78 (.09)*
ADHD	.87 (.06)	.88 (.05)	2.41 (6) *.88*	0.33 (.56)	0.21 (.65)
MDD	.77 (.06)	.82 (.07)	6.46 (7) *.49*	0.65 (.42)	0.19 (.67)
GAD	.76 (.06)	.76 (.08)	5.59 (8) *.69*	0.03 (.87)	0.02 (.90)
SAD	.45 (.13)	.51 (.09)	12.27 (8) *.72*	0.22 (.64)	0.13 (.72)
Youth Identified					
Psychiatric Disorder					
CD	.41 (.08)	.74 (.07)	2.53 (7) *.92*	8.06 (.01)†	4.33 (.04)†
CD-ODD	.57 (.09)	.70 (.08)	2.68 (6) *.85*	2.88 (.09)*	1.40 (.24)
ADHD	.54 (.14)	.65 (.12)	3.99 (6) *.68*	0.42 (.52)	2.08 (.15)
MDD	.61 (.08)	.66 (.08)	4.80 (6) *.57*	0.47 (.49)	0.14 (.71)
GAD	.58 (.09)	.41 (.12)	7.32 (6) *.29*	3.73 (.05)*	0.14 (.70)
SAD	.38 (.16)	.16 (.14)	9.03 (8) *.34*	5.65 (.02)†	0.03 (.87)

Wald χ^2 (1df): estimated loss of fit associated with constraining the unstandardized *b* coefficients and residual variance to be equal for the *MINI-KID-P* and *OCHS-EBS*. †$P<0.05$; *$P\geq0.05$, ≤0.10
Note: ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; CD-ODD: conduct disorder or oppositional defiant disorder; GAD: generalized anxiety disorder; MDD: major depressive disorder; *MINI-KID-P:* Mini International Neuropsychiatric Interview for Children and Adolescents Parent Version; *OCHS-EBS:* Ontario Child Health Study Emotional Behavioural Scales; SAD: separation anxiety disorder.

Table 4. SEM regressions of parent latent variable measures of disorder (*MINI-KID-P* and *OCHS-EBS*) on latent factor predictors after aligning the *OCHS-EBS* thresholds for classifying disorder with the *MINI-KID-P*.

	β coefficient (se)		Model fit $\chi^2_{(df)}$	Wald χ^2 (*P-value*)	
	$\beta_{MINI-KID-P}$	$\beta_{OCHS-EBS}$	*P-value*	*b*	Residual
Covariate					
Disorder					
Mental Health					
Need/Impairment					
CD	.73 (.08)	.83 (.08)	13.32 (11) .*14*	0.80 (.37)	1.32 (.25)
ADHD	.76 (.07)	.83 (.08)	13.02 (11) .*29*	0.15 (.69)	0.79 (.38)
GAD	.79 (.05)	.78 (.06)	10.17 (11) .*52*	1.52 (.22)	0.17 (.68)
Diagnostic Groupings					
CD-ODD	.81 (.06)	.71 (.07)	2.72 (6) .*84*	1.03 (.31)	4.24 (.04)†
Youth Identified					
Psychiatric Disorder					
CD	.37 (.12)	.72 (.10)	4.66 (7) .*70*	4.67 (.03)†	3.23 (.07)*
CD-ODD	.57 (.09)	.65 (.08)	11.85 (7) .*11*	0.96 (.33)	0.64 (.42)
GAD	.58 (.09)	.56 (.10)	1.55 (6) .*96*	0.57 (.45)	0.51 (.48)
SAD	.38 (.16)	.21 (.18)	8.12 (8) .*42*	1.73 (.19)	0.26 (.61)

Wald χ^2 (1df): estimated loss of fit associated with constraining the unstandardized *b* coefficients and residual variances to be equal for the *MINI-KID-P* and *OCHS-EBS*. †*P*<0.05; **P*≥0.05, ≤0.10

Note: ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; CD-ODD: conduct disorder or oppositional defiant disorder; GAD: generalized anxiety disorder; MDD: major depressive disorder; *MINI-KID-P:* Mini International Neuropsychiatric Interview for Children and Adolescents Parent Version; *OCHS-EBS:* Ontario Child Health Study Emotional Behavioural Scales; SAD: separation anxiety disorder.
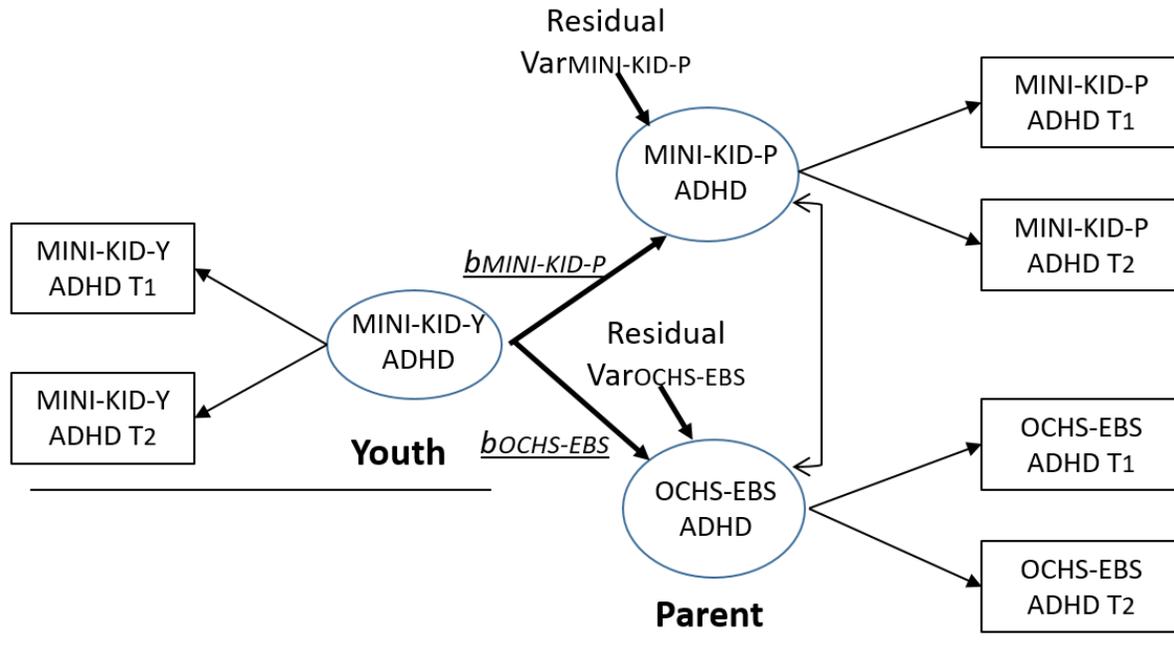
Figure 1. Structural equation model (SEM) comparing the strength of association of youth identified ADHD based on the *MINI-KID* (left side) with parent identified ADHD based on the *MINI-KID-P* and the *OCHS-EBS* (right side). Statistical comparison of this association is based on the unstandardized b coefficients ($b_{MINI-KID-P}$ versus $b_{OCHS-EBS}$) and differences in residual variance ($Var_{MINI-KID-P}$ versus $Var_{OCHS-EBS}$) – the bolded arrows. The double-headed arrow represents the covariance between the parent *MINI-KID-P* ADHD and parent *OCHS-EBS* ADHD. The other arrows represent the unique/error variance associated with the indicator variables.

Note: ADHD: attention deficit hyperactivity disorder; *MINI-KID-P:* Mini International Neuropsychiatric Interview for Children and Adolescents Parent Version; *MINI-KID-Y:* Mini International Neuropsychiatric Interview for Children and Adolescents Youth Version; *OCHS-EBS:* Ontario Child Health Study Emotional Behavioural Scales