Systematic Integration of Multi-Informant Externalizing Ratings in Clinical Settings

Madison Aitken,[1,2,3] André Plamondon,[4] John Krzeczkowski,[5] Hali Kil,[6] & Brendan F. Andrade[1,2]

[1] Centre for Addiction and Mental Health, Toronto, Canada

[2] Department of Psychiatry, University of Toronto, Canada

[3] Department of Psychology, York University, Toronto, Canada

[4] Département des fondements et pratiques en éducation, Université Laval, Québec, Canada

[5] Department of Health Sciences, Brock University, Saint Catherine's, Canada

[6] Department of Psychology, Simon Fraser University, Burnaby, Canada

Note: This is the accepted version of the manuscript, which is now published in *Research on Child and Adolescent Psychopathology*, doi: 10.1007/s10802-023-01119-z
https://link.springer.com/article/10.1007/s10802-023-01119-z

Corresponding Author:
Madison Aitken
madison.aitken@camh.ca
ORCID: 0000-0002-4921-5462

Additional ORCIDs:
André Plamondon 0000-0002-1169-2669
Hali Kil 0000-0003-4299-4552

Declarations:

The authors have no competing interests to declare that are relevant to the content of this article.

Abstract

Best practice clinical assessment of externalizing problems often necessitates collection of information from parents, youth themselves, and teachers. The present study tested the predictive validity of a psychometrically-driven scoring procedure to integrate multi-informant, dimensional ratings of externalizing problems. Participants were 2264 clinic-referred youth ages 6-18. Parents, teachers, and youth completed questionnaire ratings of externalizing problems (hyperactivity-inattention, conduct problems, and oppositionality-defiance) prior to an initial clinical appointment. The predictive validity of simple (highest informant rating; and all informant ratings separately) and more complex (latent S-1 bifactor model with specific informant factors; and moderated nonlinear factor analysis accounting for child age and sex) methods of informant integration was tested in predicting impairment, comorbidity, and number of clinical encounters. A simple model, in which all informant ratings were included, showed the best predictive validity across outcomes, performing as well or better than the use of the highest informant ratings or more complex latent variable models. The addition of child age and sex as moderators in the factor model did not improve predictive validity. Each informant (parent, teacher, and youth) contributes important information to the prediction of clinically-relevant outcomes. There is insufficient evidence at present to suggest that complex latent variable models should be favored over simpler models that preserve each informant's ratings.

Keywords: externalizing; informants; questionnaire ratings; youth

Systematic Integration of Multi-Informant Externalizing Ratings in Clinical Settings

Externalizing problems, including hyperactivity-inattention, conduct problems, and oppositionality-defiance, are among the most common reasons youth are referred for mental health services (Johnston & Burke, 2020). Increasing recognition of the limitations of categorical diagnoses has prompted a shift to using dimensional, transdiagnostic models of psychopathology in research and clinical practice (Kotov et al., 2017; Lahey et al., 2021). However, the use of dimensional measures, often in the form of questionnaires, leaves clinicians with a large amount of information to integrate. This is especially true for assessing externalizing problems in youth given that best practice clinical assessment often necessitates collection of information from parents, youth themselves, and teachers (De Los Reyes et al., 2023; Mash & Barkley, 2010; NICE, 2017). Multi-informant ratings are important for externalizing problems because youth's behaviour varies across contexts and interaction partners (De Los Reyes et al., 2009). What is missing is a method to guide clinicians on how to integrate multiple, often discrepant, reports of youth's emotional-behavioural functioning.

Using multi-informant ratings is challenging in part because agreement across informants is low, making it unclear how to reconcile discrepant ratings (De Los Reyes & Kazdin, 2005; Martel, Markon, et al., 2017). In a meta-analysis of over 300 studies of child and adolescent samples, the average correlation across informants was only low-to-moderate (r = .28; De Los Reyes et al., 2015). These low-to-moderate associations may be explained by differences in youth's behavior across settings and by informant biases and/or measurement error (De Los Reyes et al., 2013, 2015). For example, a teacher's perspective of a child's behaviour may be partly influenced by the structure of the classroom environment and social context, whereas a parent's perception may be partly influenced by their own mental health, their relationship with

the child, the home environment and other family-based factors (De Los Reyes & Epkins, 2023; De Los Reyes & Kazdin, 2005). Differences across informants may thus provide meaningful information about youth's functioning in various settings (De Los Reyes et al., 2022; Rettew et al., 2011).

Despite the importance of multi-informant ratings, clinicians may hold views about the utility of various informants' ratings that lead them to prioritize information from a single informant (De Los Reyes et al., 2022; Loeber et al., 1990; Marsh et al., 2020). For example, clinicians tend to agree with parent-endorsed goals more than youth-endorsed goals with reference to the youth's behaviour, and to make judgments about treatment response largely based on parent report (Hawley & Weisz, 2003; Marsh et al., 2020). Incorporation of multi-informant ratings may enhance care planning by providing a more accurate picture of youth externalizing problems (De Los Reyes et al., 2023).

At present, there is a lack of systematic procedures for combining information from multiple informants, leading to inconsistent practices across clinicians (De Los Reyes & Kazdin, 2005; Martel, Markon, et al., 2017). In the absence of systematic procedures for integrating multi-informant ratings, clinicians often use procedures that may alter the validity of measures (De Los Reyes & Epkins, 2023). For example, clinicians may take the highest rating across informants, use combinatorial algorithms that consider a symptom present if one or all informants endorse it, or create a composite score, such as by averaging informant ratings (Beidas et al., 2015; De Los Reyes & Epkins, 2023). With these procedures, the unique information provided about youth's behaviour in different contexts is obscured (De Los Reyes & Epkins, 2023). Integration of informant ratings relevant to the context and the reporter's

perspective may be needed to guide clinical decision making about youth externalizing problems (Beidas et al., 2015; De Los Reyes et al., 2022).

Youth sociodemographic factors, such as gender and age, may also be associated with differences in informant ratings of externalizing problems. For example, parents report more behaviour problems among boys than girls (Bajeux et al., 2018; van der Meer et al., 2008), but externalizing problems in girls are more likely to be identified by parents only, not teachers (Rettew et al., 2011). In addition, hyperactivity symptoms related to verbal social interaction are more commonly endorsed by parents and teachers in girls, whereas symptoms involving motor activity are endorsed more in boys (DuPaul et al., 2020). These differences may in part reflect informant expectations for behaviour based on gender norms. Developmental differences in the nature of externalizing behaviour may also make certain behaviours more observable to one informant than another (Lahey & Schwab-Stone, 2000). For example, oppositionality is more common in younger children, and may be readily observable to teachers and parents. In contrast, property offences are more common in older youth (Lahey & Schwab-Stone, 2000), which likely occur when parents and teachers are not present. Consideration of sociodemographic factors further complicates the meaningful integration of multi-informant data.

The use of latent variable models, in which each informant's rating is used as an indicator of a latent construct, has been suggested as a way to integrate information from multiple informants (Holmbeck, 2002; Kraemer et al., 2003; van Dulmen & Egeland, 2011). Moreover, analytic approaches, such as moderated nonlinear factor analysis have been developed, which can account for differential item functioning across groups (Curran et al., 2014; Kush et al., 2023). Applying these psychometrically driven models to multi-informant ratings of externalizing may result in better predictive validity of the resulting scores (Bauer et al., 2013).

In order to work towards development of an evidence-based and more systematic approach to integrating multi-informant data on youth externalizing problems, the present study tested the predictive validity of a psychometrically-driven scoring procedure to integrate multi-informant scores from a widely used, dimensional measure of youth psychopathology. We also determined whether predictive validity improves when accounting for youth age and sex. We expected that increasingly detailed models containing more information would have better predictive validity; however, our analysis was largely exploratory.

Method

Participants

Participants were 2264 youth ages 6 to 18 referred for outpatient mental health services at a large, publicly funded hospital in Canada. Participants were drawn from a larger sample (N = 3165) and were included in the present study if they were aged 6 to 18 years, had mental health questionnaire ratings from one or more informants (parent, teacher, or self), and had one or more demographic variables used in the present analyses (age, sex). A comparison of participants with missing and available data is presented in the Supplementary Materials. Most youth identified their ethnicity as White (63%), with 14% identifying as mixed ethnic background, 6% as East or Southeast Asian, 6% Black, 4% Latin American, 3% South Asian, 3% Middle Eastern, 1% Indigenous, 0.5% Indian-Caribbean and the remaining 0.2% another ethnicity. Parent informants were 81% mothers, 16% fathers, and 3% other relation, such as step-, foster, or grandparent.

Measures

Achenbach System of Empirically Based Assessment. Youth emotional and behavioural problems were assessed using the school-age forms from the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), including the 113-item

parent-reported Child Behavior Checklist (CBCL), the 113-item Teacher Report Form (TRF), and the 112-item Youth Self-Report (YSR) for youth ages 11-18. Each uses a 3-point Likert scale to indicate the presence of youth behaviors over the past 6 months (2 months for the TRF), whereby 0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true. Higher scores indicate more behaviour problems. The ASEBA scales have demonstrated reliability and validity as measures of youth psychopathology (Achenbach & Rescorla, 2001). *T*-scores from the Attention Problems, Rule Breaking, and Aggressive Behavior subscales were used in the present study as indicators of externalizing problems (see Supplementary Materials for rationale).

Youth Impairment. Youth impairment was used as an indicator of clinical severity (dependent variable), as measured by parent report on the 13-item Columbia Impairment Scale (CIS; Bird et al., 1993). The CIS assesses youth impairment across four areas, which include interpersonal relations, broad psychopathological domains, functioning in job or schoolwork and use of leisure time. Items are rated from 0 (no problem) to 4 (very bad problem). Item ratings are summed to create a measure of total impairment, with higher scores indicating greater impairment. The scale has demonstrated excellent reliability and validity (Bird et al., 1996). Internal consistency in the present sample was acceptable, $\alpha = 82$.

Presence of Psychiatric Comorbidity. The presence of comorbid psychiatric diagnoses was used as an indicator of clinical severity. Information was obtained from each youth's health record. Comorbidity was considered present (coded as 1) if youth were diagnosed with multiple psychiatric conditions at a single encounter or with different psychiatric conditions during different clinical encounters over time. If the youth received only one psychiatric diagnosis

and/or was given the same diagnosis at every encounter, they were classified as having no evidence of comorbidity (coded as 0).

Clinical Encounters. The number of all clinical encounters (across outpatient, inpatient, emergency visits or other encounters at our mental health hospital) was calculated for each youth and used as an indicator of clinical severity. All inpatient and emergency encounters were mental health related, but it is possible that a small number of participants with more than one encounter had an outpatient encounter that was not solely mental health-oriented (e.g., one-time consultation with a pediatrician; <1% of total sample). All but one participant with data available for the present study had one or more clinical encounters. The modal number of encounters was 1 and the data were strongly positively skewed. We therefore created an ordinal variable, consisting of 1, 2, 3, or ≥ 4 clinical encounters.

Procedure

A link to complete the questionnaires was sent to the parent and/or youth by email prior to the youth's first encounter with outpatient mental health services as a component of the initial clinical assessment. Self-report questionnaires were only sent if youth were age 13 or older. The present study uses data from June 20, 2016 to July 31, 2020. Demographic (date of birth, sex) and clinical information (clinical encounters, comorbidity) was gathered from the youth's health record. The project was approved by the hospital Research Ethics Board.

Analyses

We examined descriptive statistics for study variables. We then tested increasingly detailed models of externalizing scores to determine whether additional information improves the model's predictive validity. Specifically, we tested five models predicting three dependent variables: parent-rated impairment, presence of comorbidity, and number of clinical encounters.

Analyses were carried out in Mplus 8.2 (Muthén & Muthén, 2017) using the robust maximum likelihood estimator and the Mplus Automation package in R (Hallquist & Wiley, 2018). Missing data were handled with full information maximum likelihood estimation.

In Model 0A, a single overall Externalizing score was used, calculated by selecting the highest rating among parent, teacher, and youth self-report for each syndrome (Aggressive Behavior, Rule-Breaking, Attention Problems), then averaging these highest syndrome scores. In Model 0B, three overall Externalizing scores were used (one per informant; scores correlated). Scores were created for each informant (parent, teacher, self) by averaging *T*-scores within informant across the three syndrome scales (Aggressive Behavior, Rule-Breaking, Attention Problems).

In Model 1, we used three scores based on saved factor scores from an S-1 bifactor model consisting of a latent Externalizing factor (Aggressive Behavior, Rule-Breaking, Attention Problems) with parent as the reference informant (see Figure 1a), and orthogonal latent Teacher-Report, and Self-Report factors. Parent report was used as the reference because parent ratings of youth externalizing are widely used, likely to be available in a clinical context, and predict clinically useful outcomes longitudinally (Ferdinand et al., 2007; Holmbeck, 2002).

In Model 2, three scores were also used based on the same S-1 bifactor structure, but the latent Externalizing, Teacher-Report, and Self-Report factors were regressed on youth characteristics (age, sex) because covariates can theoretically improve the accuracy of factor scores (Curran et al., 2018). Thus, Model 2 allowed us to determine whether accounting for youth characteristics improves predictive validity.

In Model 3, we added moderation of the individual syndrome (Aggressive Behavior, Rule-Breaking; Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints) intercepts and

factor loadings by the above youth characteristics using moderated nonlinear factor analysis following the steps outlined by Kush and colleagues (Kush et al., 2023; see online Supplementary Materials for details). Model 3 allowed us to determine whether accounting for differential item functioning based on youth characteristics improves predictive validity.

Models were compared based on: 1) total variance in the dependent variable explained (R^2); 2) effect size estimates for the independent variables (standardized regression coefficients, β); and 3) 95% confidence intervals (CIs) for the standardized regression coefficients. To account for the different number of predictors in some models, we also examined adjusted β values (see Table S3); however, as the pattern of results remained the same, we retained the standard β values, for which confidence intervals were available. Youth age and sex were used as covariates in all regression models.

Results

Descriptive Analysis

There were 2095 youth with parent ratings, 1192 with teacher ratings, and 1015 with self-report ratings. Participants with and without parent, teacher, or self-report ratings differed in terms of some demographic variables (age, sex, ethnicity) and the severity of ratings provided by available informants (see Supplementary Materials for details). Descriptive statistics for study variables are presented in Table 1, and correlations between continuous variables are presented in Table S1. Fit indices for the S-1 bifactor model were as follows: comparative fit index (CFI) = 0.88, root mean square error of approximation (RMSEA) = .11.

Predictive Validity

Impairment. R^2 values for the five models are presented in Figure 2. In predicting parent-rated impairment, the R^2 value was lowest for Model 0A, in which Externalizing scores

were created by averaging the highest scores across informants for each syndrome scale. The R^2 value was highest for Model 0B, in which an overall Externalizing score for each of parents, teachers, and youth was entered in the model. Models 1 (S-1 bifactor model), 2 and 3 (in which youth age and sex were accounted for), performed comparably to one another and less well than Model 0B in terms of predicting variance in parent-rated impairment based on non-overlapping 95% CIs for the R^2 values.

Standardized beta coefficients for each of the models are presented in Figure 3a. Coefficients for the overall Externalizing factor tended to be larger than for the informant-specific (teacher, self) factors. Beta coefficients within factor types were comparable, with overlapping confidence intervals across simple and more complex models, including models accounting for youth characteristics in estimating the factors, with one exception: Model 2 (S-1 bifactor model with factors regressed on youth age and sex) had a larger coefficient for the overall Externalizing factor than did the Externalizing score in the simple models (0A and 0B). In examining the significance of predictors of impairment in the model with the highest predictive validity based on R^2 values (Model 0B), parent-rated and teacher-rated Externalizing each significantly predicted parent-rated impairment, though self-rated Externalizing did not (see Table 2; see Supplementary Materials for results of Models 0A and 1-3).

Comorbidity. R^2 values for models predicting comorbidity were much smaller than for parent-rated impairment (see Figure 2). Again, the more complex models, in which youth characteristics were accounted for (Models 2 and 3), did not result in improved predictive validity compared to simple models (Models 0A and 0B) as the 95% CIs all overlapped.

Confidence intervals for standardized beta coefficients within factor types (overall externalizing vs. self- or teacher-report; see Figure 3b) overlapped, suggesting that more

complex models accounting for youth characteristics in estimating factor scores did not result in increased predictive validity of the factors. In the model with the highest predictive validity based on R^2 values (Model 0B), higher parent- and teacher-rated Externalizing, but not self-rated Externalizing, significantly predicted the presence of comorbidity in youth's clinical profile, (see Table 2; see Supplementary Materials for results of Models 0A and 1-3).

Number of Clinical Encounters. R^2 values for models predicting the number of clinical encounters were much smaller than for parent-rated impairment but comparable to models predicting comorbidity (see Figure 2). The more complex models did not result in better prediction of clinical encounters than did simpler models, as evidenced by overlapping 95% CIs for the R^2 values.

Standardized beta coefficients (see Figure 3c) again indicated little improvement in predictive validity with increasing model complexity. In the model with the highest predictive validity based on R^2 values (Model 0B), higher parent- and self-rated, but not teacher-rated, Externalizing significantly predicted more clinical encounters (see Table 2; see supplementary results for full results of Models 0A and 1-3).

Discussion

The present study tested whether the use of a psychometrically-driven approach to integrating multi-informant ratings of youth externalizing problems improved predictive validity compared to simpler approaches. Externalizing problems were broadly defined and included hyperactivity-inattention, conduct problems, and oppositionality-defiance, consistent with current dimensional models of psychopathology (Kotov et al., 2017; Martel, Pan, et al., 2017). We also tested whether accounting for moderating effects of youth demographics would improve the predictive validity of externalizing factor scores. We found that a simple method using an

overall externalizing score for each of parents, teachers, and youth performed as well or better than more complex latent variable models. The addition of child age and sex as moderators of factor loadings and intercepts in the latent variable model also did not improve the predictive validity of the resulting factor scores.

Traditional latent variable approaches treat variance unique to each informant as error (Holmbeck, 2002). In contrast, we used an S-1 bifactor model, which allowed us to model both shared variance across informants and variance specific to teacher and youth informants (Burns et al., 2020). Despite this complex modeling, we found that the use of a latent variable approach to integrating multi-informant ratings did not improve the prediction of clinically-relevant outcomes compared to simpler approaches. Our results are similar to recent findings in a large community sample of youth that psychopathology factors derived through bifactor modeling may not have better predictive validity than factors from simple correlated factor models (Watts et al., 2019). While it may be conceptually argued that we should separate common vs. informant-specific variance, it will be important to identify models that demonstrate their added-value.

Our results supported the use of a simple approach to using multi-informant ratings, in which parent, teacher, and youth self-report ratings of externalizing are simultaneously considered. Importantly, this model showed better predictive validity than another simple approach using the highest scores across parent, teacher, and youth report, particularly in predicting parent-rated impairment. Taking the highest score among informants' continuous ratings is equivalent to the "or" rule, in which a symptom is considered present if endorsed by one informant (Gizer et al., 2008). The small number of studies comparing the predictive validity of various methods of combining informant ratings have generally found similar results (Gizer et

al., 2008; Offord et al., 1996). The present results are therefore consistent with the large body of evidence that each informant contributes important information (Charamut et al., 2022; De Los Reyes et al., 2022; Rettew et al., 2011). Using algorithms that collapse information across informants, such as the "or" rule, may therefore omit clinically relevant information.

Our results also support the importance of youth self-report of externalizing problems. Self-reported externalizing problems predicted clinical encounters, whereas teacher-rated externalizing problems did not. Although collateral reports of externalizing problems in youth are often favored (Loeber et al., 1990; Marsh et al., 2020), findings from the present study show that self-report adds clinically important information, potentially about youth's perceived need for mental health support.

We also found that accounting for moderation of factor scores, item intercepts, and item loadings by child age and sex did not improve the predictive validity of factor scores. Perhaps using item-level data would lead to different findings, although such models would prove more complex to estimate for scales with large numbers of items.

Limitations and Future Directions

The results of the present study should be considered with the following limitations in mind. First, our clinic-referred sample is not representative of the general population and results may in part reflect referral patterns. Second, our dependent variables were primarily cross-sectional (though comorbidity and number of encounters may have incorporated future clinical information) and either rated by parents (impairment) or had the potential to be influenced by parents (comorbid diagnoses; number of clinical encounters). Future studies using longitudinal methods and a broader range of clinical outcomes would provide stronger information. Third, we had limited information available on youth characteristics (including youth ethnicity due to high

rates of missing data on this variable in the electronic medical records). We also did not examine whether the various informant integration models tested were more or less useful across different ethnocultural or gender identity groups. Future studies incorporating additional factors, including gender, socioeconomic status, and ethnocultural identity, are needed given evidence of their potential effects on informant ratings (Dirks et al., 2011; Lau et al., 2004; Offord et al., 1996).

Future studies could test alternative strategies to model multi-informant data, including traditional bifactor models, given that some studies have found better model fit in bifactor compared to S-1 bifactor models (Watts et al., 2022). Further studies will be useful to test different scoring strategies to incorporate multi-informant data, ideally using longitudinal, multi-informant data to test their predictive validity. Similar analyses on other domains of psychopathology, including internalizing, are also needed. Finally, further work is needed to develop tools that clinicians can use to consider information from multiple informants in a systematic manner.

Conclusion

With a shift toward dimensional measures being used in both research and practice (Krueger et al., 2018), there is a need for an evidence-informed method of considering ratings from multiple informants. Multi-informant ratings of youth externalizing problems are important in predicting clinically-relevant markers of severity, including impairment, comorbidity, and number of clinical encounters. Researchers and clinicians should collect and consider multi-informant ratings given that each predicts clinically relevant outcomes related to severity of a given youth's mental health needs. In integrating multi-informant externalizing ratings, clinicians and researchers can use easy-to-implement strategies, at least until more sophisticated approaches can demonstrate their added value.

Supplementary Materials:

https://static-content.springer.com/esm/art%3A10.1007%2Fs10802-023-01119-z/MediaObjects/10802_2023_1119_MOESM1_ESM.docx

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Bajeux, E., Klemanski, D. H., Husky, M., Leray, E., Chan Chee, C., Shojaei, T., Fermanian, C., & Kovess-Masfety, V. (2018). Factors associated with parent-child discrepancies in reports of mental health disorders in young children. *Child Psychiatry & Human Development*, *49*(6), 1003–1010. https://doi.org/10.1007/s10578-018-0815-7

Bauer, D. J., Howard, A. L., Baldasaro, R. E., Curran, P. J., Hussong, A. M., Chassin, L., & Zucker, R. A. (2013). A trifactor model for integrating ratings across multiple informants. *Psychological Methods*, *18*(4), 475–493. https://doi.org/10.1037/a0032475

Beidas, R. S., Stewart, R. E., Walsh, L., Lucas, S., Downey, M. M., Jackson, K., Fernandez, T., & Mandell, D. S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and Behavioral Practice*, *22*(1), 5–19. https://doi.org/10.1016/j.cbpra.2014.02.002

Bird, H. R., Andrews, H., Schwab-Stone, M., Goodman, S., Dulcan, M., Richters, J., Rubio-Stipec, M., Moore, R. E., Chiang, P., Hoven, C., Canino, G., Fisher, P. W., & Gould, M. S. (1996). Global measures of impairment for epidemiological and clinical use with children and adolescents. *International Journal of Methods in Psychiatric Research*, *6*(4), 295–307. https://doi.org/10.1002/(SICI)1234-988X(199612)6:4<295::AID-MPR173>3.3.CO;2-5

Bird, H. R., Shaffer, D., Fisher, P., & Gould, M. S. (1993). The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research*, *3*(3), 167–176.

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020). Application of the bifactor S-1 model to multisource ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. *Journal of Abnormal Child Psychology*, *48*(7), 881–894. https://doi.org/10.1007/s10802-019-00608-4

Charamut, N. R., Racz, S. J., Wang, M., & De Los Reyes, A. (2022). Integrating multi-informant reports of youth mental health: A construct validation test of Kraemer and colleagues' (2003) Satellite Model. *Frontiers in Psychology*, *13*, 911629. https://doi.org/10.3389/fpsyg.2022.911629

Curran, P. J., Cole, V. T., Bauer, D. J., Rothenberg, W. A., & Hussong, A. M. (2018). Recovering predictor–criterion relations using covariate-informed factor score estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 860–875. https://doi.org/10.1080/10705511.2018.1473773

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*, *49*(3), 214–231. https://doi.org/10.1080/00273171.2014.889594

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Deborah, A. G., Burgers, D. E., Rabinowitz, J., Reyes, A. D. L., Augenstein, T. M., Wang, M., Thomas, S. A., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037.a0038498

De Los Reyes, A., & Epkins, C. C. (2023). Introduction to the special issue: A dozen years of demonstrating that informant discrepancies are more than measurement error—Toward

guidelines for integrating data from multi-informant assessments of youth mental health. *Journal of Clinical Child & Adolescent Psychology*, *52*(1), 1–18. https://doi.org/10.1080/15374416.2022.2158843

De Los Reyes, A., Epkins, C. C., Asmundson, G. J. G., Augenstein, T. M., Becker, K. D., Becker, S. P., Bonadio, F. T., Borelli, J. L., Boyd, R. C., Bradshaw, C. P., Burns, G. L., Casale, G., Causadias, J. M., Cha, C. B., Chorpita, B. F., Cohen, J. R., Comer, J. S., Crowell, S. E., Dirks, M. A., … Youngstrom, E. A. (2023). Editorial statement about JCCAP's 2023 special issue on informant discrepancies in youth mental health assessments: Observations, guidelines, and future directions grounded in 60 years of research. *Journal of Clinical Child & Adolescent Psychology*, *52*(1), 147–158. https://doi.org/10.1080/15374416.2022.2158842

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*(5), 637–642. https://doi.org/10.1007/s10802-009-9307-3

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483

De Los Reyes, A., Talbott, E., Power, T. J., Michel, J. J., Cook, C. R., Racz, S. J., & Fitzpatrick, O. (2022). The needs-to-goals gap: How informant discrepancies in youth mental health assessments impact service delivery. *Clinical Psychology Review*, *92*, 102114. https://doi.org/10.1016/j.cpr.2021.102114

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617

Dirks, M. A., Boyle, M. H., & Georgiades, K. (2011). Psychological symptoms in youth and later socioeconomic functioning: Do associations vary by informant? *Journal of Clinical Child and Adolescent Psychology*, *40*(1), 10–22. https://doi.org/10.1080/15374416.2011.533403

DuPaul, G. J., Fu, Q., Anastopoulos, A. D., Reid, R., & Power, T. J. (2020). ADHD Parent and Teacher Symptom Ratings: Differential Item Functioning across Gender, Age, Race, and Ethnicity. *Journal of Abnormal Child Psychology*, *48*(5), 679–691. https://doi.org/10.1007/s10802-020-00618-7

Ferdinand, R. F., Van Der Ende, J., & Verhulst, F. C. (2007). Parent-teacher disagreement regarding psychopathology in children: A risk factor for adverse outcome? *Acta Psychiatrica Scandinavica*, *115*(1), 48–55. https://doi.org/10.1111/j.1600-0447.2006.00843.x

Gizer, I. R., Waldman, I. D., Abramowitz, A., Barr, C. L., Feng, Y., Wigg, K. G., Misener, V. L., & Rowe, D. C. (2008). Relations between multi-informant assessments of ADHD symptoms, DAT1, and DRD4. *Journal of Abnormal Psychology*, *117*(4), 869–880. https://doi.org/10.1037/a0013297

Hallquist, M. N., & Wiley, J. F. (2018). *MplusAutomation*: An R Package for Facilitating Large-Scale Latent Variable Analyses in M*plus*. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Hawley, K. M., & Weisz, J. R. (2003). Child, parent and therapist (dis)agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. *Journal of Consulting and Clinical Psychology*, *71*(1), 62–70. https://doi.org/10.1037/0022-006X.71.1.62

Holmbeck, G. N. (2002). Collecting and managing multisource and multimethod fata in studies of pediatric populations. *Journal of Pediatric Psychology*, *27*(1), 5–18. https://doi.org/10.1093/jpepsy/27.1.5

Johnston, O. G., & Burke, J. D. (2020). Parental problem recognition and help-seeking for disruptive behavior disorders. *The Journal of Behavioral Health Services & Research*, *47*(1), 146–163. https://doi.org/10.1007/s11414-018-09648-y

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kraemer, H. C., Measelle, J., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. (2003). A new approach to integrating data from multiple sources in psychiatric assessments and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, *160*(September), 1566–1577. https://doi.org/10.1176/appi.ajp.160.9.1566

Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., Simms, L. J., Widiger, T. A., Achenbach, T. M., Bach, B., Bagby, R. M., Bornovalova, M. A., Carpenter, W. T., Morey, L. C., Mullins-sweatt, S. N., Ormel, J., Patalay, P., Patrick, C.

J., & Pincus, A. L. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry*, *17*(3).

Kush, J. M., Masyn, K. E., Amin-Esmaeili, M., Susukida, R., Wilcox, H. C., & Musci, R. J. (2023). Utilizing moderated non-linear factor analysis models for integrative data analysis: A tutorial. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(1), 149–164. https://doi.org/10.1080/10705511.2022.2070753

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry*, *20*(1), 57–63. https://doi.org/10.1002/wps.20824

Lahey, B. B., & Schwab-Stone, M. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology*, *109*(3), 488–503.

Lau, A. S., Garland, A. F., Yeh, M., Mccabe, K. M., Wood, P. A., & Hough, R. L. (2004). Race/ethnicity and inter-informant agreement in assessing adolescent psychopathology. *Journal of Emotional and Behavioral Disorders*, *12*(3), 145–156. https://doi.org/10.1177/10634266040120030201

Loeber, R., Green, S. M. S., & Lahey, B. B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, *19*(2), 136–143. https://doi.org/10.1207/s15374424jccp1902

Marsh, J. K., Zeveney, A. S., & De Los Reyes, A. (2020). Informant discrepancies in judgments about change during mental health treatments. *Clinical Psychological Science*, *8*(2), 318–332. https://doi.org/10.1177/2167702619894905

Martel, M. M., Markon, K., & Smith, G. T. (2017). Research Review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(2), 116–128. https://doi.org/10.1111/jcpp.12611

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., Manfro, G. G., Miguel, E. C., Paus, T., Bressan, R. A., Rohde, L. A., & Salum, G. A. (2017). A general psychopathology factor (P factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, *126*(1), 137–148. https://doi.org/10.1037/abn0000205

Mash, E. J., & Barkley, R. A. (2010). *Assessment of childhood disorders* (4th ed). Guilford.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide*.

NICE. (2017). *Antisocial behaviour and conduct disorders in children and young people: Recognition and management*. https://www.nice.org.uk/guidance/CG158

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(8), 1078–1085. https://doi.org/10.1097/00004583-199608000-00019

Rettew, D. C., van Oort, F. V. A., Verhulst, F. C., Buitelaar, J. K., Ormel, J., Hartman, C. A., Veenstra, R., Althoff, R. R., & Hudziak, J. J. (2011). When parent and teacher ratings don't agree: The Tracking Adolescents' Individual Lives Survey (TRAILS). *Journal of Child and Adolescent Psychopharmacology*, *21*(5), 389–397. https://doi.org/10.1089/cap.2010.0153

van der Meer, M., Dixon, A., & Rose, D. (2008). Parent and child agreement on reports of problem behaviour obtained from a screening questionnaire, the SDQ. *European Child & Adolescent Psychiatry*, *17*(8), 491–497. https://doi.org/10.1007/s00787-008-0691-y

van Dulmen, M. H. M., & Egeland, B. (2011). Analyzing multiple informant data on child and adolescent behavior problems: Predictive validity and comparison of aggregation procedures. *International Journal of Behavioral Development*, *35*(1), 84–92. https://doi.org/10.1177/0165025410392112

Watts, A. L., Makol, B. A., Palumbo, I. M., De Los Reyes, A., Olino, T. M., Latzman, R. D., DeYoung, C. G., Wood, P. K., & Sher, K. J. (2022). How robust is the p factor? Using multitrait-multimethod modeling to inform the meaning of general factors of youth psychopathology. *Clinical Psychological Science*, *10*(4), 640–661. https://doi.org/10.1177/21677026211055170

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, *7*(6), 1285–1303. https://doi.org/10.1177/2167702619855035

Table 1

Descriptive Statistics

Variable	% or *M(SD)*	Range
Age	12.5(3.6)	6.0 - 18.2
Sex		
Female	46	
Male	54	
Impairment	22.65 (9.59)	0 - 52
Comorbidity	62	
Clinical encounters		
1	36	
2	33	
3	15	
≥4	16	
Parent-rated		
Attention problems	63.35 (10.02)	50 - 100
Rule breaking	59.99 (8.62)	50 - 91
Aggressive behavior	63.39 (11.04)	50 - 100
Teacher-rated		
Attention problems	60.19 (8.54)	50 - 96
Rule breaking	59.40 (8.37)	50 - 84
Aggressive behavior	62.48 (11.71)	50 - 98
Self-rated		
Attention problems	65.76 (11.47)	50 - 100
Rule breaking	58.87 (8.34)	50 - 100
Aggressive behavior	58.72 (8.84)	50 - 100

Table 2

Regression Model 0B (Individual Parent, Teacher, and Youth Externalizing Ratings as Independent Variables)

Independent Variable	Impairment			Comorbidity			Number of Encounters		
	β	*SE*	*p*	β	*SE*	*p*	β	*SE*	*p*
Parent-rated externalizing	**0.67**	0.02	<.001	**0.16**	0.05	.002	**0.10**	0.03	.003
Teacher-rated externalizing	**0.11**	0.03	<.001	**0.13**	0.06	.025	0.01	0.04	.872
Self-rated externalizing	0.04	0.03	.162	0.05	0.06	.438	**0.12**	0.04	.004
Age	**0.15**	0.02	<.001	0.06	0.05	.172	-0.03	0.03	.383
Sex (male)	**0.06**	0.02	<.001	0.02	0.04	.619	**0.06**	0.02	.016

Note. Bold values are significant at $p < .05$

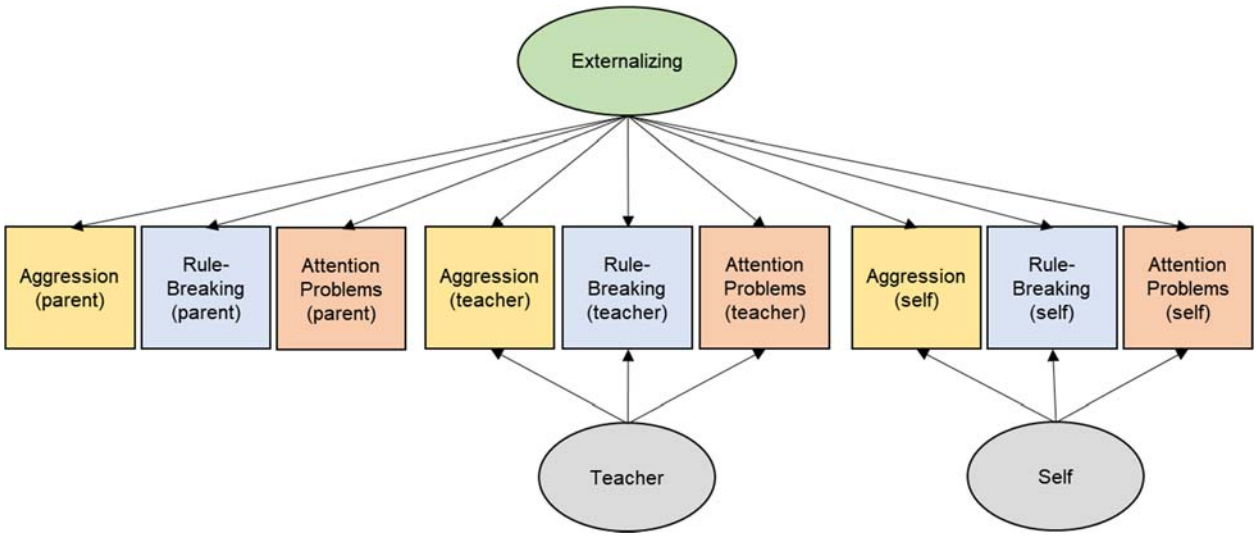

Figure 1. S-1 bifactor model

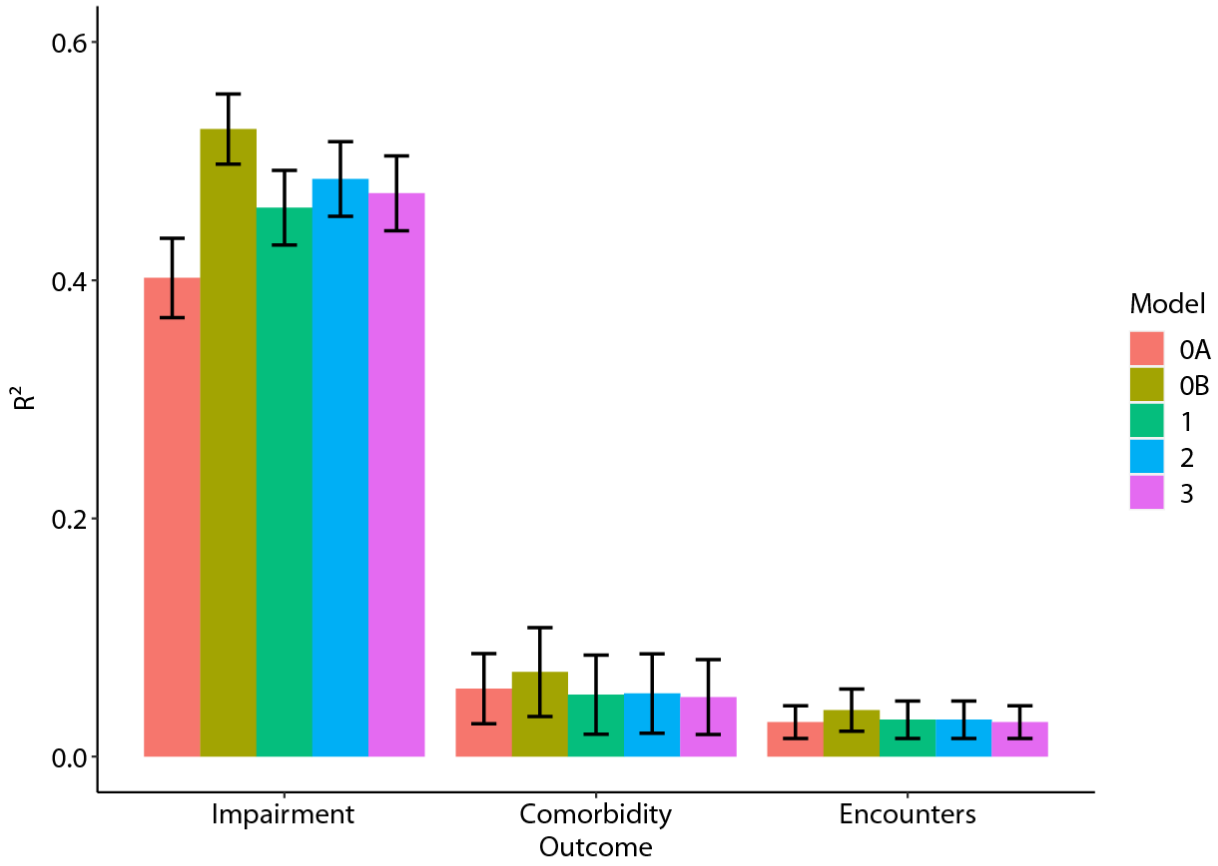

Figure 2. R^2 values for the five regression models

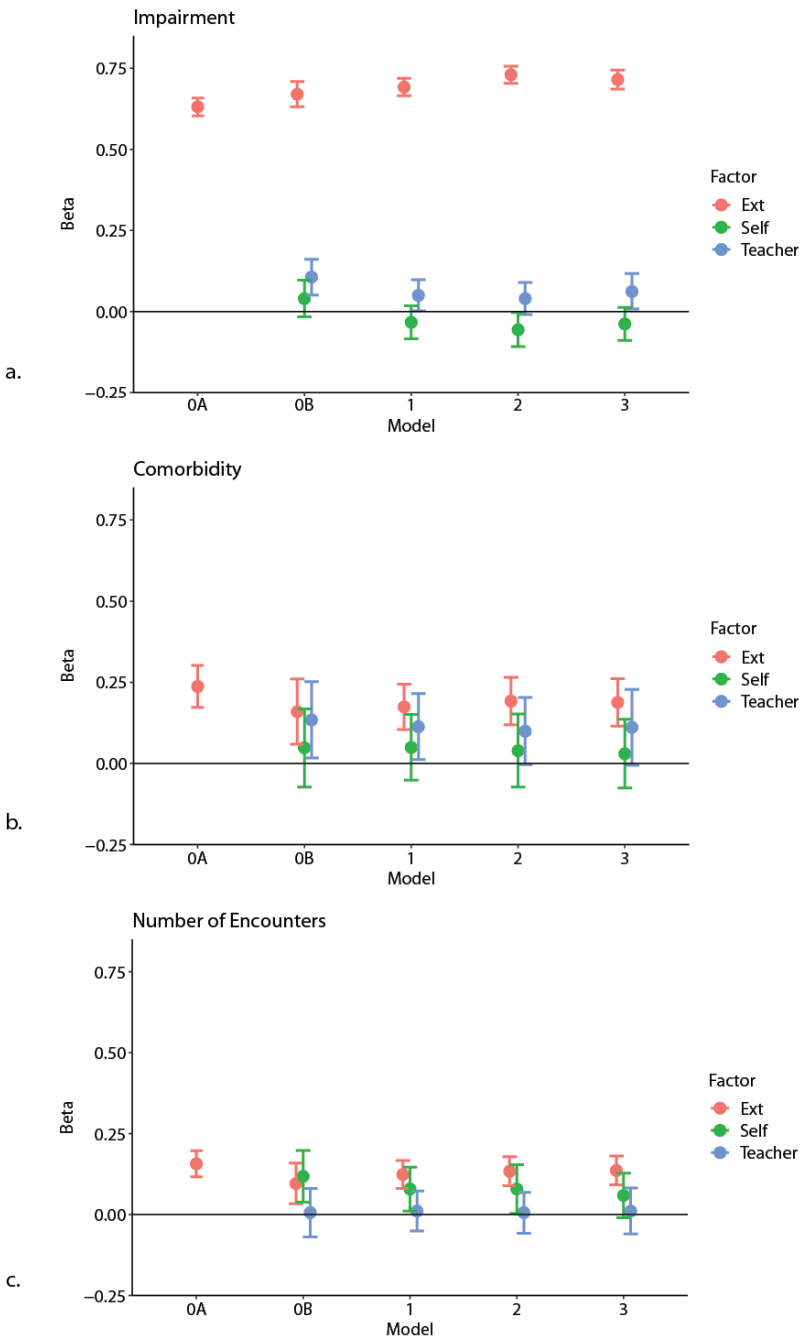

Figure 3. Standardized beta coefficients. Note that the Ext coefficient in Model 0B reflects parent externalizing ratings only.