Developmental Models of Depression, Externalizing Problems, and Self-regulatory Processes: Integrated Data Analysis Across Four Longitudinal Studies of Youth

Kelsey E. Magee[1] · Arin Connell[1] · Alison E. Hipwell[2,3] · Daniel Shaw[3] · Erika Westling[4] · Kate Keenan[5] · Elizabeth Stormshak[6] · Thao Ha[7] · Stephanie Stepp[2,3]

Abstract

Integrative data analysis (IDA) was used to derive developmental models of depression, externalizing problems, and self-regulatory processes in three prevention trials of the Family Check-Up and one longitudinal, community-based study of girls over a 10-year span covering early to late adolescence ($N = 4{,}773$; 74.9% female, 41.7% white). We used moderated nonlinear factor analysis to create harmonized scores based on all available items for a given participant in the pooled dataset while accounting for potential differences in both the latent factor and the individual items as a function of observed covariates. We also conducted latent growth model analyses to examine developmental trajectories of risk. Results indicated a bidirectional relationship between depression and externalizing problems, with greater baseline externalizing problems and depression predicting growth in inhibitory control difficulties. Furthermore, initial level of inhibitory control difficulties was associated with growth in depression. We did not, however, find a relationship between early inhibitory control difficulties and growth in externalizing problems. This work illustrates the utility of IDA techniques to harmonize data across multiple studies to identify risk factors for the development of depression and externalizing problems that can be targeted by prevention efforts.

Introduction

The transition from childhood to adolescence is often marked by significant psychosocial, emotional, and neurobiological changes that can result in mood and behavior

✉ Kelsey E. Magee
  kem151@case.edu

[1] Department of Psychological Sciences, Case Western Reserve University, Cleveland, USA

[2] Department of Psychiatry, University of Pittsburgh, Pittsburgh, USA

[3] Department of Psychology, University of Pittsburgh, Pittsburgh, USA

[4] Oregon Research Institute, Eugene, USA

[5] Department of Psychiatry, University of Chicago, Chicago, USA

[6] University of Oregon, College of Education, Eugene, USA

[7] Department of Psychology, Arizona State University, Phoenix, USA

difficulties (Auerbach et al., 2017; Ingoldsby et al., 2006). For instance, depression becomes far more prevalent during adolescence and is associated with a multitude of deleterious outcomes, including suicidal ideation and attempts, often with tragic consequences (McLeod et al., 2016). Additionally, externalizing problems, including a broad array of behavior problems, also increase during adolescence and frequently co-occur with depression (Ingoldsby et al., 2006; Wolff & Ollendick, 2011). The onset of emotional and behavioral problems during childhood and adolescence is often associated with chronic, lifelong impairment.

Poor self-regulatory abilities in youth are associated with heightened risk for impulsivity, depression, and externalizing problems (Connell et al., 2019). Indeed, several models focused on self-regulatory processes describe the pathways through which externalizing problems may lead to depression, and vice versa. The cumulative failure model contends that externalizing behavior problems may result in escalating difficulties in social and academic functioning that subsequently lead to depression (Gardner et al., 2008;

Patterson & Stoolmiller, 1991). Another theory posits that depression in youth may be associated with heightened risk for externalizing problems via inhibitory control difficulties. For instance, the negative emotions that often characterize depression (e.g., sadness, irritability) impair executive functioning necessary for self-regulation (see Curci et al., 2013; Chester et al., 2016). Specifically, when youth struggle to disengage from negative, repetitive thought patterns associated with rumination, they may experience an intense aversion to these negative affective experiences (e.g., negative urgency). Negative urgency, a facet of impulsivity that predicts problematic externalizing behaviors above and beyond other features of impulsivity, may motivate vulnerable youth to quickly diminish the strength of these emotions by impulsively engaging in deviant and/or aggressive behaviors (Chester et al., 2016; Connell et al., 2019).

Such developmental models have begun to provide evidence for the dynamic interplay of how emotional and behavioral vulnerabilities interact to increase risk for depression and externalizing problems; however, long-term, longitudinal studies of the antecedents of risk remain sparse (Kerr et al., 2012). There is a pressing need for research to identify potentially modifiable developmental processes that may either amplify or attenuate risk for these difficulties across adolescence that can be targeted by prevention efforts. The present study involves novel secondary data analyses, employing integrated data analysis (IDA) procedures to examine developmental models of depression, externalizing problems, and self-regulatory processes across one community-based study of girls, and three large scale prevention trials with boys and girls. All four studies include rich data collected during overlapping assessment points across critical developmental spans for the onset of depression and externalizing problems. Linking these studies through IDA increases statistical power and provides a deeper understanding of how these constructs develop over time (Bainter & Curran, 2015).

Integrative Data Analysis

The use of IDA capitalizes upon between-study heterogeneity to promote a better understanding of differences in findings across studies and to probe meaningful sources of between-study variability that may contribute to, and inform theories about, key psychological phenomenon (Curran & Hussong, 2009). In the current study, the psychometric constructs related to risk for depression and externalizing problems were substantially broadened by incorporating diagnostic data. When items varied across studies, IDA techniques were used to create comparable estimates of constructs across datasets that contain at least some common items, coupled with unique items that appear in only one or

a subset of datasets (Curran et al., 2014). In this way, reliable and valid commensurate measures were obtained and used in subsequent pooled analyses across the four datasets (Curran et al., 2014).

In IDA, latent constructs, such as "depression severity" and "externalizing problems," can be better measured using several items (vs. a single item) to capture the underlying latent construct (Gottfredson et al., 2019). Specifically, we employed a moderated nonlinear factor analysis (MNLFA) approach, conducted with an automated R package (aMNLFA; Gottfredson et al., 2019). Using MNLFA, we created harmonized scores based on all available items for a given participant in the pooled dataset while accounting for potential differences in both the latent factor and the individual items as a function of observed covariates (Bauer & Hussong, 2009; Curran et al., 2014). The MNLFA approach simultaneously tests whether a measure is invariant across important factors, such as age and gender, with respect to factor means and variances, as well as item intercepts and factor loadings (Curran et al., 2014). For instance, MNLFA can directly test whether a continuously distributed covariate, such as age, is systematically associated with higher levels of depression (the factor mean) and greater variability of depression (the factor variance) while avoiding the need to create discrete, dichotomous groups based on age (e.g., old vs. young; Curran et al., 2014). Finally, MNLFA can test whether specific items on a depression scale are differentially endorsed (the item intercepts) or more strongly indicative (the factor loadings) of underlying depression when age is measured on a continuum (Curran et al., 2014). Such tests are particularly important in IDA, and are not possible using standard item response theory, differential item functioning, or confirmatory factor analysis approaches alone (Curran et al., 2014; Gottfredson et al., 2019).

Current Study

The current study examined developmental trajectories of depression, externalizing problems, and self-regulatory processes in a harmonized dataset that includes data collected across a 10-year span, with overlapping assessment points in early through late adolescence (ages 11 to 20). To maximize content similarity across study waves, DSM-5 criteria for major depressive disorder, oppositional defiant disorder (ODD), and conduct disorder (CD) were used to identify items that correspond with symptoms of depression and externalizing problems (See supplemental Tables 1–3 and the Supplemental Data Map for a list of available items across studies).

Consistent with the theory of negative urgency (Chester et al., 2016; Connell et al., 2019) that posits that youth

may engage in externalizing behaviors in an attempt to diminish the strength of negative emotions, we hypothesized that youth depression would predict growth in later externalizing problems. Conversely, based on the cumulative failure model and findings that youth with conduct problems may experience social and developmental failures associated with depression risk (Gardner et al., 2008; Patterson & Stoolmiller, 1991), we hypothesized that externalizing problems would predict growth in depression. Although self-regulatory problems in youth are associated with heightened risk for a range of difficulties across development, including depression and externalizing problems, long-term longitudinal models elucidating intervening pathways are limited. Our recent work suggests that improvements in self-regulation across childhood are associated with reductions in depression and suicide risk in adolescence (Connell et al., 2019). The current study aims to expand upon these findings by examining how deficits in self-regulatory processes are associated with depression and externalizing problems, exploring the strength and directionality of these relationships over time. Since none of the four studies in the current analyses was set up to test race and ethnicity as moderators of the development of emotional and behavioral problems, we did not include race and ethnicity in our hypotheses. However, we did control for minority status as a potential demographic confound, as data on race were available in all four studies. IDA techniques are advantageous to the analyses in the current study because they incorporate a longer longitudinal time span (versus analyses in any one trial) based on sample heterogeneity in ages and by harmonizing scores across diverse measurement strategies than were used in individual samples over time.

Methods

Description of Primary Study Samples

Three of the four samples in the current analyses are prevention trials of the Family Check-Up (FCU) intervention program, which was designed to reduce the development of externalizing problems and substance use by improving parenting practices and enhancing family relationships, although cross-over effects on depression and suicide risk have been observed across trials (e.g., Connell et al., 2016, 2018, 2021). The FCU follows an adaptive intervention framework, in which intervention targets and doses are tailored to the individual needs of families, promoting increased family engagement and more efficient use of resources (Collins et al., 2004). Individuals in the control condition of the FCU did not receive any access to the intervention and were essentially treated as participants in

a longitudinal study, in that families were simply asked to complete study assessments over time.

Early Steps

This prevention trial sample includes 731 primary caregiver–child dyads. Families with children between the ages of 2 years 0 months and 2 years 11 months were recruited between 2002 and 2003 from WIC programs around Pittsburgh, Pennsylvania, Eugene, Oregon, and Charlottesville, Virginia (see Dishion et al., 2008). Families completed comprehensive assessments at 10 study waves, from ages 2 through age 16 years. Retention was above 80% for most assessment points, including an 81% retention rate at the most recent completed waves (age 14–16), with 79% retention at the lowest assessment point (age 7.5).

Project Alliance 1 (PAL 1)

This prevention trial sample includes 998 adolescents and their families, recruited from three middle schools within a predominantly low resourced metropolitan community in the northwestern United States, when youth were in sixth grade (see Dishion et al., 2012). Parents of all sixth-grade students in two cohorts were approached for participation, and 90% consented to participate. Families completed comprehensive assessments at 9 study waves, from age 11 years to 28–30 years. Retention was above 80% for most assessment points, with 75.6% retention at the final assessment point.

Project Alliance 2 (PAL 2)

This prevention trial sample includes 593 families recruited when their child was in sixth grade, from three predominantly low resourced middle schools in the northwestern United States (see Stormshak et al., 2011). Parents of all sixth-grade students in two cohorts were approached for participation, and 76% consented to participate. Families completed comprehensive assessments at 7 study waves, from age 11 years to age 23 years. Retention was above 80% for most assessment points, with 78% of participants completing early adult assessments.

Pittsburgh Girls Study (PGS)

This prospective study includes 2450 girls recruited between the ages of 5 and 8 years in 2000–2001 via stratified, random household sampling, with over-sampling of households in low resourced neighborhoods (see Keenan et al., 2010a, b). Retention was above 85% for all assessment points, with 87.3% of participants completing early adult assessments.

Demographic Questionnaires

Demographic questionnaires were administered, yielding information regarding gender and race/ethnicity.

Questionnaire-Based Reports of Depression

Early Steps

Youth completed the Child Depression Inventory (CDI; Kovacs, 1992), a self-report measure of depressive symptoms, at ages 14 and 16.

PAL 1 Youth completed the CDI (Kovacs, 1992) at ages 11–13, and the Brief Symptom Inventory (BSI; Derogatis & Fitzpatrick, 1982), at age 16, which includes a depression symptom scale. At age 18, youth also completed the Youth Self Report (YSR) version of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000). The CBCL/YSR includes a scale measuring symptoms of depression.

PAL 2 Youth completed a depression symptom checklist (DSC; Klostermann et al., 2016) at ages 11–14, including 14 items reflecting past-month severity of symptoms of depression. Youth also completed the CDI (Kovacs, 1992) at ages 11–13 as part of the FCU assessment. At age 20, young adults completed the Adult Self-Report (ASR; Achenbach & Rescorla, 2000), which is the adult self-report version of the YSR/CBCL.

PGS Youth completed the Child Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 1994) during the age 11–12 assessments, the Adolescent Symptom Inventory-4 (ASI-4; Gadow & Sprafkin, 1998) during the age 13–17 assessments, and the Adult Self-Report Inventory-4 (ASRI; Gadow et al., 1994) during the age 18–20 assessments.

Questionnaire-Based Reports of Externalizing Problems

Early Steps

Youth completed the Self-Report of Delinquency (SRD; Elliott et al., 1985) scale at age 10.5, 14, and 16, including items that represent the frequency of "acts for which juveniles could be arrested (e.g., theft, disorderly conduct)" during the past year.

PAL 1 Youth completed the YSR/ASR at ages 11–13 and 18, which includes scales measuring externalizing problems. Externalizing problems were measured at age 11–14 via self-report of nine items (Student School Survey). Youth

also completed the Student Self Check (SSC), a self-report measure of behavior problems at age 11–13.

PAL 2 Externalizing problems were measured via the Student School Survey at age 11–14 and the ASR at age 20.

PGS Youth completed the CSI-4 (Gadow & Sprafkin, 1994) during the age 12 assessment, the ASI-4 (Gadow & Sprafkin, 1998) during the age 13–17 assessments, and the ASRI-4 (Gadow et al., 1994) during the age 18–20 assessments.

Questionnaire-Based Reports of Self-Regulatory Processes

Early Steps, PAL 1, and PAL 2

Youth completed the Early Adolescent Temperament Questionnaire-Revised (EATQ-R; Capaldi & Rothbart, 1992), assessing temperament and self-regulation, as part of the FCU yearly assessment (Early Steps, ages 10.5, 14 and 16; PAL 1, age 17; PAL 2, age 20). Items from the Inhibitory Control subscale of EATQ-R were included in the current analyses. Items reflecting "positive" inhibitory control (e.g., "I can stick with my plans and goals") were reverse scored for consistency. Higher values indicate poorer self-regulation. Additionally, youth in PAL 1 completed the YSR at ages 11–13. Youth in PAL 1 and PAL 2 completed the ASR (PAL 1, age 18; PAL 2, age 20).

PGS Youth completed the Social Skills Rating System-Child (SSRS-C; Gresham & Elliot, 1990) as part of the yearly assessment at ages 11–18. Three items from the SSRS-C were mapped onto items of the EATQ-R reflecting inhibitory control. As was done in the prevention trials, items were reverse scored for consistency and higher values indicate poorer self-regulation.

Analytic Plan

Data harmonization analyses employed MNLFA (Hussong et al., 2013), which involves an iterative series of analyses in which possible covariate effects on item difficulty and discrimination, as well as on latent variable means and variances are examined sequentially. Youth reports of depression, externalizing problems, and inhibitory control were analyzed using an R-based aMNLFA package (Gottfredson et al., 2019) to iteratively format Mplus command files. MNLFA analyses were conducted in Mplus 8.4 (Muthén & Muthén, 1998–2017).

Following Curran et al. (2014), a single time-point of data for each participant was randomly selected to generate a calibration sample for establishing measurement properties to avoid problems with non-independence of longitudinal data.

Using the calibration sample, an iterative series of analyses was conducted to examine invariance across covariates with respect to factor means, variances, and item intercepts/factor loadings to obtain valid item parameter estimates adjusting for DIF. Categorical covariates were included using an effects-coding approach, as follows: gender (0 = male, 1 = female), minority status (0 = white, 1 = racial minority), intervention-assignment (0 = control, 1 = intervention), and study (with three orthogonal contrasts, comparing the PGS with PAL 1, PAL 2, and Early Steps; Early Steps with PAL 1 and PAL 2; and PAL 1 with PAL 2).

We examined potential covariate differences in overall mean and variance for which individual items reflected symptom severity (i.e., factor loadings) and in the likelihood of item-endorsement across levels of the covariate for individuals at the same level of severity (i.e., item intercepts). Significant covariate effects were then compiled into a larger "simultaneous" model, which was used to estimate latent variable model parameters (e.g., factor loadings, item intercepts, variances) for items reflecting the latent variable while controlling for these covariate effects. Parameter estimates from the "simultaneous" model are then fixed in a subsequent "scoring model," including only significant effects that survive correction, which then uses the full longitudinal dataset to generate point estimates of the latent variable of interest based on observed covariates and item responses. This final scoring model was then used to generate estimates using the longitudinal dataset, which were then used in subsequent latent growth model (LGM) analyses (see Duncan & Duncan, 2009). We controlled for minority status in the LGM analyses with three orthogonal contrasts, comparing white vs. all other minority groups, Black vs. Latinx and all others, and Latinx vs. all other minority groups.

Missing Data

There are several aspects of missing data relevant to these analyses. First, as in any study with long-term follow-up,

attrition has occurred in each sample. However, as described in more detail in the "Methods" section, retention was above 80% for most assessment waves for all studies. We have found that attrition patterns are consistent with "missing at random" assumptions in prior analyses related to depression, externalizing problems, and self-regulatory processes in each of the studies. Additionally, as in any study, some participants failed to complete individual measures at each study wave, although such missing data rates were relatively low. Finally, at some study waves, planned missingness was incorporated into assessments. Specifically, in the first 3 waves of PAL 1, for example, high-risk youth were identified by teacher reports, and were asked to complete additional measures of depression (see Connell & Dishion, 2008).

Results

Descriptive Statistics

The aggregated sample included data from 4773 participants. The number of participants in the intervention versus control groups and additional demographic information are shown in Table 1. Race and ethnicity were assessed separately in the PGS. None of the prevention trials collected data on race and ethnicity separately, but rather included Hispanic/Latinx as one option in a list of self-identified race/ethnicity responses that participants could select.

Data Harmonization

Depression

MNLFA analyses were conducted for youth reports of depressive symptoms. Following a comprehensive examination of available items and guided by diagnostic criteria, as well as the example of Curran and colleagues (2014), we mapped items from all available measures onto 17 symptoms

Table 1 Demographic characteristics

Study	Baseline (years)	Endpoint (years)	N	Female (%)	European American (%)	African American (%)	Latinx (%)	Multiracial (%)
ES	2	16	364 control	49.5	46.7	26.6	13.2	9.9
			367 intervention	49.6	46.6	28.6	13.6	9.8
PAL 1	11	30	498 control	48.2	42	28.3	7.4	11.8
			500 intervention	46.4	42.8	30	6.2	10.6
PAL 2	11	23	207 control	49.3	36.7	14.5	18.4	19.8
			386 intervention	48.2	35.8	15.5	17.9	18.9
PGS	5–8	24–27	2,450 total	100	41.2	52.9	2	4.9

ES, Early Steps, prevention trial; *PAL 1*, Project Alliance 1, prevention trial; *PAL 2*, Project Alliance 2, prevention trial; *PGS*, Pittsburgh Girls Study

of depression. Item responses were dichotomized to represent the endorsement of a given symptom across measures (0 = no, 1 = yes). Estimates for significant factor loadings for each symptom, covariate effects on item intercepts (reflecting differences in item endorsement across levels of the covariate for those at the same level of depression severity), and differential item functioning in factor loadings are shown in Supplemental Table 4. Covariate effects on the factor mean parameter (reflecting mean level differences in depression across covariate levels) are shown in Supplemental Table 5. There were no significant covariate effects on the factor variance.

Externalizing Problems

Items from all available measures were mapped onto 23 externalizing symptoms included in the diagnostic criteria for ODD and CD, and item responses were dichotomized to represent the endorsement of a given symptom across measures (0 = no, 1 = yes). The following symptoms were not included in analyses because they were not assessed consistently across studies: *angry or resentful*, *deliberately annoys others*, and *spiteful or vindictive*. Furthermore, two symptoms, *has been physically cruel to animals* and *has broken into someone else's house, building, or car*, had non-significant factor loadings. Significant factor loadings for each symptom, covariate effects on item intercepts, and differential item functioning in factor loadings are shown in Supplemental Table 6. Covariate effects on the factor mean parameter are shown in Supplemental Table 5. There were no significant covariate effects on the factor variance.

Self-Regulatory Processes

Items from all available measures were mapped onto the 17 items of the EATQ-R reflecting inhibitory control and were dichotomized to represent the endorsement of a given item across measures (0 = no, 1 = yes) according to a logical harmonization approach. Items that were reversed scored are indicated with an asterisk. Higher values indicate poorer inhibitory control. Significant factor loadings for each symptom and differential item functioning in factor loadings are shown in Supplemental Table 7. Covariate effects on the factor mean parameter are presented in Supplemental Table 5. There were no significant covariate effects on the factor variance.

Latent Growth Model (LGM)

We conducted a single LGM analysis examining growth in depression, externalizing problems, and inhibitory control difficulties over time. We first conducted a series of analyses to generate an acceptable-fitting model with several

covariates, including gender, minority status, intervention status, and study level contrasts (see supplemental Table 8 for covariate effects). The final model with covariates provided acceptable fit to the data by most criteria ($\chi 2 = 7595.37$, df = 583, $p < 0.05$, CFI = 0.906/TLI = 0.900, RMSEA = 0.05; see Fig. 1 for LGM with paths connecting latent intercepts and slopes). Random assignment to the intervention was not significantly associated with changes in the slope parameter for depression, externalizing problems, and inhibitory control; thus, we included data from participants from both the intervention and control conditions in the LGM analyses.

### Associations Between Constructs over Time

Growth in depression was significantly associated with greater baseline externalizing problems (Estimate = 0.055, SE = 0.00, $p = 0.000$) and greater baseline inhibitory control problems (Estimate = 0.029, SE = 0.01, $p = 0.003$). Growth in externalizing problems was significantly associated with greater baseline depression (Estimate = 0.037, SE = 0.00, $p = 0.000$), but was not significantly associated with baseline inhibitory control. Finally, growth in inhibitory control difficulties was significantly associated with greater baseline externalizing problems (Estimate = 0.010, SE = 0.00, $p = 0.009$) and greater baseline depression (Estimate = 0.017, SE = 0.00, $p = 0.000$).

Discussion

The IDA approach in the current study represents a comprehensive examination of the development of depression, externalizing problems, and self-regulatory processes over a 10-year span across early through late adolescence in three randomized prevention trials of the FCU and one community-based longitudinal study of girls. Aggregating multiple datasets increases sample heterogeneity and allows for the incorporation of multiple measurement approaches for a broader and more rigorous examination of key constructs across a long developmental span (Curran & Hussong, 2009). We will first discuss the results of the harmonization analyses, highlighting how they inform our understanding of the constructs of interest, followed by a discussion of the hypothesis testing analyses, highlighting ways that researchers can make novel use of the estimates from MNLFA.

Harmonization Analyses

We examined study-level differences directly relevant to the feasibility of data aggregation, as they allow for the identification and adjustment of differences in measurement properties across samples. Specifically, we examined
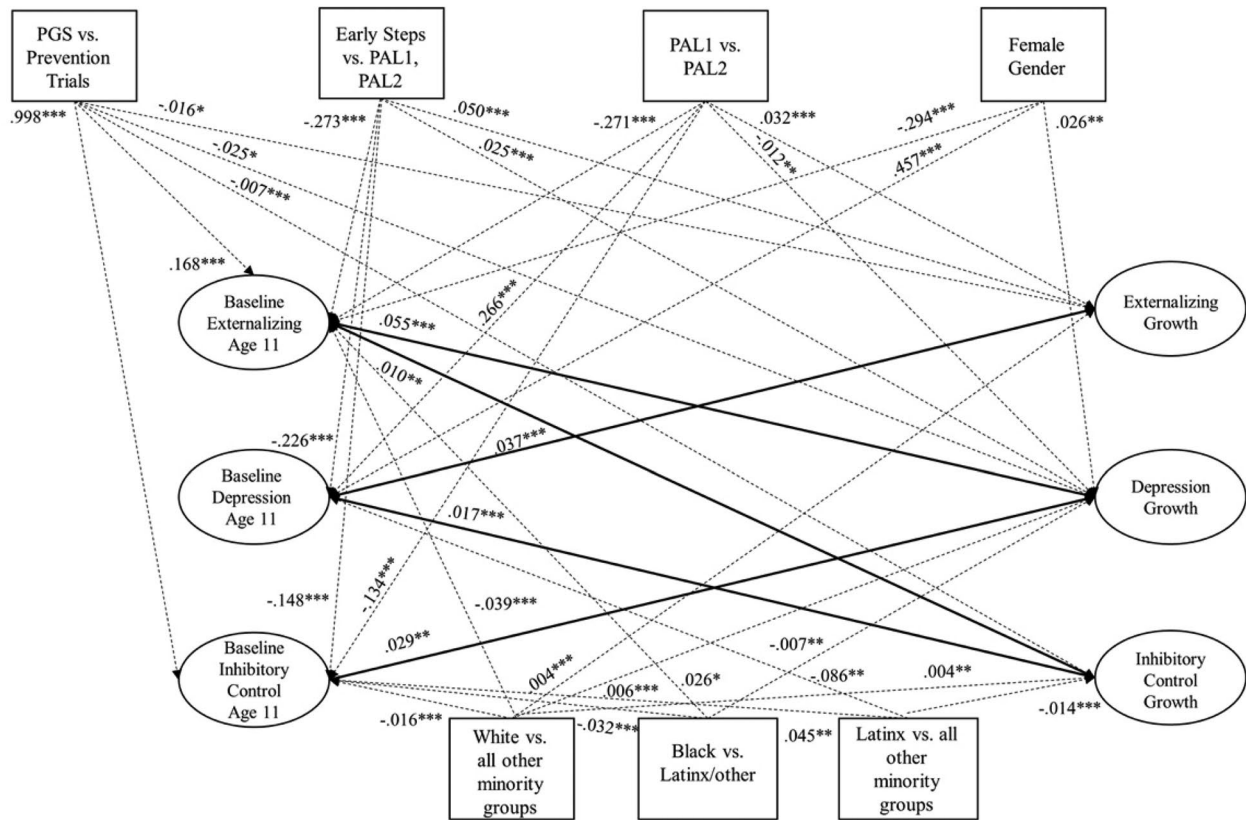
Note. *p < .05. **p < .01. ***p < .001.*

Fig. 1 Final latent growth model with study level contrasts and gender as covariates. Significant standardized estimates and are reported, and significance levels are as follows

whether a given symptom was endorsed more in one sample versus others at the same level of the latent variable or was more strongly related to the latent variable. We found, for instance, that the factor means for depression, externalizing problems, and self-regulatory processes were higher in the PGS relative to the three prevention trials, and such study-level differences may result from the original sampling methodology. Additionally, we examined mean level differences in the latent variables across several demographic and study-based covariates, which may identify specific subgroups of risk based on gender, intervention assignment, and age at assessment. We found, for instance, that the factor mean for externalizing problems was lower for girls than boys. This finding is consistent with the broader literature on gender differences in externalizing problems in childhood and adolescence (e.g., Ingoldsby et al., 2006) and our previous work indicating that adolescent onset CD in girls is rare (Keenan et al., 2010a, b), but adds to the literature by demonstrating the power of IDA techniques to investigate group differences across multiple studies. Furthermore, we examined covariate effects on item intercepts, which reflect a youth's tendency to differentially endorse a given item

across levels of the covariate at the same underlying level of the latent variable. For depression, for instance, older age at assessment was associated with greater rates of endorsement of *anhedonia*, *fatigue*, *worthlessness*, *suicidal ideation*, and *suicide attempt*, indicating that at the same underlying level of depression, older youth are more likely to endorse these symptoms. Finally, we examined DIF in factor loadings, which allowed us to determine if symptoms are more indicative of the underlying construct in one group versus another, which is particularly helpful for improving our understanding of assessment approaches. For instance, the factor loadings for three symptoms of depression, *anhedonia*, *appetite disturbance*, and *decreased energy*, were more positive with increasing age, suggesting that youth reports of these symptoms are stronger indicators of depression in later adolescence. We controlled for these covariate effects in the LGMs so they did not cloud our hypothesis testing analyses.

LGM Analyses

Integrating data with overlapping ages allows for the examination of important constructs across longer developmental

periods, within a shorter amount of time than can be done within one study (Bainter & Curran, 2015). In the Early Steps study, for instance, child report data were collected at ages 10.5, 14, and 16. However, the other three studies collected data through late adolescence, allowing for examination of a longer developmental span than would be possible in separate analyses within the Early Steps sample. Furthermore, the PGS includes only girls, and so incorporating mixed gender samples from the prevention trials increases sample heterogeneity and permits broader tests of developmental pathways across genders that are not possible in the original sample. Finally, IDA facilitates item-level analyses across datasets when items from different measures have been used across trials. In the present study, we created scores for externalizing symptoms using 21 items assessing ODD and CD. Specific items, such as *touchy or easily annoyed* were more strongly related to the underlying level of externalizing in females, and we were able to control for this heterotypic continuity in the LGM analyses.

The results of the LGM analyses in the current harmonized dataset indicate a bidirectional relationship between depression and externalizing problems across a 10-year span, with early externalizing problems predicting growth in depression, and early depression predicting growth in externalizing problems. Our findings add to the literature, as the few studies that examined internalizing symptoms as child-related risk factors for later externalizing problems have relied on parent and teacher report of symptoms (Jarrett et al., 2014) or have found that internalizing and externalizing problems do not have reciprocal relations throughout early development in community samples (Stone et al., 2015).

Our finding that early externalizing problems predict growth in depression is consistent with the dual failure model of depression, which suggests that externalizing problems are associated with psychosocial "failures," including peer rejection and academic difficulties, leading to negative self-perception and depression (Patterson & Stoolmiller, 1991; Patterson et al., 1992; Rice et al., 2017). Although we did not directly test the mediating or moderating role of such peer relational and academic challenges in the longitudinal associations between externalizing problems and depression, items in the inhibitory control subscale included self-regulatory difficulties related to peer and academic experiences. Notably, we also found that growth in depression was associated with greater baseline inhibitory control difficulties, suggesting that both self-regulatory difficulties in peer and academic domains, and externalizing problems may drive increases in depression during adolescence. This finding expands upon our previous work showing that improvements in inhibitory control across childhood were associated with prospective decreases in depression and suicidality (Connell et al., 2019). Taken together, results of the current study indicate that the

outward expression of externalizing behaviors, potentially driven by self-regulatory difficulties already evident in early adolescence (age 11), may serve as identifiable risk factors for depression across the adolescent period. Thus, prevention efforts should focus on screening for externalizing problems earlier childhood (e.g., in early childhood or at formal school entry; Lochman & Conduct Problems Prevention Research Group, 1995; Shaw et al., 2003) so that they can be targeted by family focused prevention programs (e.g., the early childhood version of the FCU) before they worsen and result in difficulties across important life domains.

Our finding that depression predicts externalizing problems is consistent with patterns of negative urgency, where youth may exhibit dysregulated behavior associated with externalizing problems as an attempt to avoid strong negative emotions that characterize depression (Chester et al., 2016). However, we did not find a relationship between early inhibitory control difficulties and externalizing problem, which is in contrast with the broader literature in this area that suggests that low levels of inhibitory control predict externalizing behavior problems in young children (Gagne et al., 2018) and college aged adults (Venables et al., 2018). Our novel findings indicate that growth in externalizing problems in the context of depression is related to emotion-driven behaviors, rather than self-regulatory difficulties across a long developmental span into late adolescence. Taken together, our results help to inform efforts to prevent externalizing problems, suggesting that prevention programs should teach youth to identify, process, and cope with emotions rather than avoid them (Modecki et al., 2017).

Finally, our finding that indicates that greater baseline externalizing problems and depression predict growth in inhibitory control difficulties underlies the importance of the implementation of preventive efforts targeting risk related to both behavior problems (e.g., dysregulation related to impulsivity) and depression (e.g., negative urgency) during childhood to ameliorate the burden of these difficulties over time.

Limitations and Future Directions

There are limitations to the current study that highlight broader challenges with IDA techniques and suggest avenues for future work. First, it may be difficult to identify studies and measures to include in an IDA project. For instance, we chose inhibitory control as the measure of self-regulatory processes given the role of inhibitory control difficulties in the development of depression and externalizing problems; however, negative emotionality would be another reasonable option. Thus, IDA projects require collaborations with the original study investigators to establish buy-in, identify measures for inclusion, and to address potential questions about study design issues that may affect analyses. It is, of course, important to include studies that have conceptually

similar data at overlapping time points to ensure that the data can be aggregated to adequately address the question of interest, while at the same time, ensuring that there are differences in the selected studies that IDA can capitalize upon. In the current study, symptoms of ODD were measured in only two studies in early and mid-adolescence, which limited our ability to test for the differential onset, course, and predictive utility of these symptoms early in development across all four studies. Additionally, as our IDA project included data from three prevention trials and one longitudinal study without an intervention component, we had to be mindful of potential effects of the prevention program on results and determine how to proceed with hypotheses testing. For instance, we chose to control for effects of the intervention in the LGM analysis. As we did not observe a significant effect of the intervention, we reported results from the entire sample. Another option would be to only include individuals in the control condition in analyses; however, this option would greatly reduce the sample size, and thus the statistical power of the results.

Second, the task of identifying commensurate measures across studies is a challenge for data synthesis that requires researchers to make decisions on how to identify the items to include in analyses. In the current study, we aggregated depression and externalizing items at the symptom level rather than including all available items, and we collapsed all responses into dichotomized scores. We believe that our choice is meaningful because it included the most overlap of items across studies that are clinically relevant to symptoms of depression and externalizing disorders endorsed at threshold or not at all. However, a potential disadvantage to our approach is that it does not address nuances that may present with the option of scale level responses. Relatedly, we mapped inhibitory control items onto the EATQ-R to include the most overlap of items across studies relevant to self-regulatory processes. However, the content of items reflecting inhibitory control varied across studies, and our approach may have excluded some important aspects of inhibitory control when items did not map onto the EATQ-R.

Finally, another challenge lies in identifying covariates to include in data harmonization and hypothesis testing analyses. In the current study, we included a basic set of covariates that were readily available to harmonize across the samples. However, potentially important additional covariates were either not assessed or were assessed inconsistently across studies. For instance, none of the four studies in the current analyses was set up to test race *and* ethnicity as moderators of the development of emotional and behavioral problems, and separate data on ethnicity were not obtained in the three prevention trials. Thus, we controlled for minority status as a potential demographic confound, but we did not include race or ethnicity in our hypotheses. Furthermore, factors such as SES or other aspects of child and family functioning may be important with respect to DIF but were not available for our analyses. Relatedly, we are currently harmonizing data to include additional variables of peer relational factors (e.g., peer deviance, peer rejection) that may serve as mediators to facilitate the examination of longitudinal mechanistic pathways between depression, externalizing problems, and self-regulatory difficulties across studies. We are interested in how these peer relational factors relate to self-regulatory processes, and how they may interact to serve as risk and/or protective mechanisms in the development of depression and externalizing problems. Such research may yield novel insights into risk factors that can be targeted by prevention efforts.

Conclusion

The current study employed IDA techniques to examine developmental models of depression, externalizing problems, and inhibitory control difficulties in three prevention trials of the FCU and one longitudinal, community-based study of girls. Our results highlight the benefits and potential challenges of utilizing IDA to synthesize data across multiple studies to examine growth in mood and behavioral difficulties across a 10-year span. Results of the harmonization analyses revealed study and covariate level differences in symptom endorsement and mean level differences in constructs of interest across covariates, emphasizing the utility of IDA to potentially identify subgroups of risk based on symptoms endorsed. Furthermore, LGM analyses in the harmonized dataset allowed us to test developmental models beyond what would be possible in the individual studies. Results of the LGM analyses provided novel insight into developmental targets for prevention and intervention programs, including the importance of targeting depression, externalizing problems, and inhibitory control difficulties during childhood to prevent growth in symptoms over time.

Declarations

Consent to Participate Not applicable (secondary data analysis).

Conflict of Interest The authors declare that they have no conflict of interest.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles* (Vol. 30). University of Vermont, Research center for children, youth, & families.

Auerbach, R. P., Stewart, J. G., & Johnson, S. L. (2017). Impulsivity and suicidality in adolescent inpatients. *Journal of Abnormal Clinical Psychology, 45*, 91–103. https://doi.org/10.1007/s10802-016-0146-8

Bainter, S. A., & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development, 16*, 1–20. https://doi.org/10.1080/15248372.2013.871721

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods, 14*, 101–125. https://doi.org/10.1037/a0015583

Capaldi, D. M., & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament measure. *The Journal of Early Adolescence, 12*, 153–173. https://doi.org/10.1177/0272431692012002002

Chester, D. S., Lynam, D. R., Milich, R., Powell, D. K., Andersen, A. H., & DeWall, N. (2016). How do negative emotions impair self-control? A neural model of negative urgency. *NeuroImage, 132*, 43–50. https://doi.org/10.1016/j.neuroimage.2016.02.024

Collins, L. M., Murphy, S. A., & Bierman, K. L. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science, 5*, 185–196. https://doi.org/10.1023/b:prev.0000037641.26017.00

Connell, A., & Dishion, T. (2008). Reducing depression among at-risk early adolescents: Three-year effects of a family-centered intervention embedded within schools. *Journal of Family Psychology, 22*, 574–585. https://doi.org/10.1037/0893-3200.22.3.574

Connell, A. M., Magee, K., Stormshak, E., Ha, T., Westling, E., Wilson, M., & Shaw, D. (2021). Long-term cross over effects of the family check-up prevention program on child and adolescent depression: Integrative data analysis of three randomized trials. *Journal of Consulting and Clinical Psychology, 89*, 773–782. https://doi.org/10.1037/ccp0000677

Connell, A. M., McKillop, H. N., & Dishion, T. J. (2016). Long-term effects of the family check-up in early adolescence on risk of suicide in early adulthood. *Suicide and Life-Threatening Behavior, 46*, S15–S22. https://doi.org/10.1111/sltb.12254

Connell, A. M., Shaw, D., Wilson, M., Danzo, S., Weaver-Krug, C., Lemery-Chalfant, K., & Dishion, T. (2019). Indirect effects of the early childhood Family Check-Up on adolescent suicide risk: The mediating role of inhibitory control. *Development and Psychopathology, 31*, 1901–1910. https://doi.org/10.1017/S0954579419000877

Connell, A. M., Stormshak, E., Dishion, T., Fosco, G., & Van Ryzin, M. (2018). The Family Check Up and adolescent depression: An examination of treatment responders and non-responders. *Prevention Science, 19*(1), 16–26. https://doi.org/10.1007%2Fs11121-015-0586-3

Curci, A., Lanciano, T., Soleti, E., & Rimé, B. (2013). Negative emotional experiences arouse rumination and affect working memory capacity. *Emotion, 13(5),* 867–880. https://doi.org/10.1037/a0032492

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*, 81–100. https://doi.org/10.1037/a0015914

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*, 214–231. https://doi.org/10.1080/00273171.2014.889594

Derogatis, L. R., & Fitzpatrick, M. (1982). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18. In M. E. Maruish (Ed.), The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults (pp. 1–41). Lawrence Erlbaum Associates Publishers.

Dishion, T. J., Ha, T., & Veronneau, M. H. (2012). An ecological analysis of the effects of deviant peer clustering on sexual promiscuity, problem behavior, and childbearing from early adolescence to adulthood: An enhancement of the life history framework. *Developmental Psychology, 48*, 703. https://doi.org/10.1037/a0027304

Dishion, T. J., Shaw, D. S., Connell, A., Gardner, F., Weaver, C., & Wilson, M. (2008). The family check-up with high-risk indigent families: Outcomes of positive parenting and problem behavior from ages 2 through 4 years. *Child Development, 79*, 1395–1414. https://doi.org/10.1111/j.1467-8624.2008.01195.x

Duncan, T. E., & Duncan, S. E. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass*, *3*(6), 979–991. https://doi.org/10.1111/j.1751-9004.2009.00224.x

Elliott, D. S., Ageton, S. S., & Huizinga, D. (1985). *Explaining Delinquency and Drug Use*. Sage.

Gadow, K. D. & Sprafkin, J. (1994). *Child Symptom Inventories manual*. Stony Brook, NY: Checkmate Plus.

Gadow, K. D. & Sprafkin, J. (1998). *Adolescent Symptom Inventory-4 screening manual*. Stony Brook, NY: Checkmate Plus.

Gadow, K. D., Sprafkin, J., & Weiss (1994). *Adult Self-Report Inventory-4 Manual*. Stony Brook, NY: Checkmate Plus.

Gagne, J. R., Chang, C., Fang, H., Spann, C., & Kwok, O. (2018). A multimethod study of inhibitory control and behavioural problems in preschoolers. *Infant and Child Development, 28*, e2115. https://doi.org/10.1002/icd.2115

Gardner, T. W., Dishion, T. J., & Connell, A. M. (2008). Adolescent self-regulation as resilience: Resistance to antisocial behavior within the deviant peer context. *Journal of Abnormal Child Psychology, 36*, 273–284. https://doi.org/10.1007/s10802-007-9176-6

Gottfredson, N. C., Cole, V. T., Giordano, M. L., Bauer, D. J., Hussong, A. M., & Ennett, S. T. (2019). Simplifying the implantation of modern scale scoring methods with an automated r package: Automated moderated nonlinear factor analysis (aMNLFA). *Addictive Behaviors, 94*, 65–73. https://doi.org/10.1016/j.addbeh.2018.10.031

Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating System manual*. American Guidance Service.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9*, 61–89. https://doi.org/10.1146/annurev-clinpsy-050212-185522

Ingoldsby, E. E., Kohl, G. O., McMahon, R. J., & Lengua, L. (2006). Conduct problems, depression symptomatology, and their co-occurring presentation in childhood as predictors of adjustment in early adolescence. *Journal of Abnormal Child Psychology, 34*, 603–621. https://doi.org/10.1007/s10802-006-9044-9

Jarrett, M., Siddiqui, S., Lochman, J., & Qu, L. (2014). Internalizing problems as a predictor of change in externalizing problems in at-risk youth. *Journal of Clinical Child & Externalizing Psychology, 43*, 27–35. https://doi.org/10.1007/s10802-006-9044-9

Keenan, K., Hipwell, A., Chung, T., Stepp, S., Stouthamer-Loeber, M., Loeber, R., & McTigue, K. (2010a). The Pittsburgh Girls Studies: Overview and initial findings. *Journal of Clinical Child & Adolescent Psychology, 39*, 506–521. https://doi.org/10.1080/15374416.2010.486320

Keenan, K., Wroblewski, K., Hipwell, A., Loeber, R., & Stouthamer-Loeber, M. (2010b). Age of onset, symptom threshold, and expansion of the nosology of conduct disorder for girls. *Journal of Abnormal Psychology*, *119*(4), 689–698. https://doi.org/10.1037/a0019346

Kerr, D. C. R., Reinke, W. M., & Eddy, L. M. (2012). Trajectories of depressive symptoms and externalizing behaviors across adolescence: Associations with histories of suicide attempt and ideation in early adulthood. *Suicide and Life-Threatening Behavior*, *43*(1), 50–66. https://doi.org/10.1111/j.1943-278x.2012.00127.x

Klostermann, S., Connell, A., & Stormshak, E. A. (2016). Gender differences in the developmental links between conduct problems and depression across early adolescence. *Journal of Research on Adolescence, 26*, 76–89. https://doi.org/10.1111/jora.12170

Kovacs, M. (1992). Children's depression inventory. North Tonawanda, NY: Multi-Health Systems.

Lochman, J. E. & Conduct Problems Prevention Research Group. (1995). Screening for child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology, 63*, 549–559. https://doi.org/10.1037/0022-006X.63.4.549

McLeod, G. F. H., Horwood, L. J., & Fergusson, D. M. (2016). Adolescent depression, adult mental health and psychosocial outcomes at 30 and 25 years. *Psychological Medicine, 46*, 1401–1412. https://doi.org/10.1017/S0033291715002950

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus User's Guide (8th ed.) Los Angeles, CA: Muthén & Muthén.

Modecki, K. L., Zimmer-Gembeck, M., & Guerra, N. (2017). Emotion regulation, coping, and decision making: Three linked skills for preventing externalizing problems in adolescence. *Child Development, 88*, 417–426. https://doi.org/10.1111/cdev.12734

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). Antisocial boys. Eugene, OR: Castalia.

Patterson, G., & Stoolmiller, M. (1991). Replications of a dual failure model for boys' depressed mood. *Journal of Consulting and Clinical Psychology, 59*, 491–498. https://doi.org/10.1037//0022-006x.59.4.491

Rice, F., Sellers, R., Hammerton, G., Eyre, O., Bevan-Jones, R., Thapar, A. K., Collishaw, S., Harold, G. T., & Thapar, A. (2017). Antecedents of new-onset major depressive disorder in children and adolescents at high familial risk. *JAMA Psychiatry, 74*, 153–160. https://doi.org/10.1001/jamapsychiatry.2016.3140

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. (2003). lead article). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*, 189–200. https://doi.org/10.1037//0012-1649.39.2.189

Stone, L. L., Otten, R., Engels, R. C. M., Kuijpers, R. C. W. M., & Janssens, J. M. A. M. (2015). Relations between internalizing and externalizing problems in early childhood. *Child & Youth Care Forum, 44*, 635–653. https://doi.org/10.1007/s10566-014-9296-4

Stormshak, E. A., Connell, A. M., Véronneau, M., Myers, M. W., Dishion, T. J., Kavanagh, K., & Caruthers, A. S. (2011). An ecological approach to promoting early adolescent mental health and social adaptation: Family-centered intervention in public middle schools. *Child Development*, 209–225. https://doi.org/10.1111/j.1467-8624.2010.01551.x

Venables, N. C., Foell, J., Yancey, J. R., Kane, M. J., Engle, R. W., & Patrick, C. J. (2018). Quantifying inhibitory control as externalizing proneness: A cross-domain model. *Clinical Psychological Science, 6*, 561–580. https://doi.org/10.1177/2167702618757690

Wolff, J. C., & Ollendick T. H. (2011). Conduct problems in youth: Phenomenology, classification, and epidemiology. In: Murrihy R., Kidman A., Ollendick T. (Eds.) Clinical Handbook of Assessing and Treating Conduct Problems in Youth. Springer: New York, NY.