



Regular Article

The general psychopathology factor (p) from adolescence to adulthood: Exploring the developmental trajectories of p using a multi-method approach

Alexandria M. Choate¹ , Marina A. Bornovalova¹, Alison E. Hipwell² , Tammy Chung³ and Stephanie D. Stepp²

¹Department of Psychology, University of South Florida, Tampa, FL, USA, ²Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA and

³Department of Psychiatry, Institute for Health, Healthcare Policy and Aging Research; Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

Abstract

Considerable attention has been directed towards studying co-occurring psychopathology through the lens of a general factor (p -factor). However, the developmental trajectory and stability of the p -factor have yet to be fully understood. The present study examined the explanatory power of dynamic mutualism theory – an alternative framework that suggests the p -factor is a product of lower-level symptom interactions that strengthen throughout development. Data were drawn from a population-based sample of girls ($N = 2450$) who reported on the severity of internalizing and externalizing problems each year from age 14 to age 21. Predictions of dynamic mutualism were tested using three distinct complementary statistical approaches including: longitudinal bifactor models, random-intercept cross-lagged panel models (RI-CLPMs), and network models. Across methods, study results document preliminary support for mutualistic processes in the development of co-occurring psychopathology (that is captured in p). Findings emphasize the importance of exploring alternative frameworks and methods for better understanding the p -factor and its development.

Keywords: adolescence; co-occurring psychopathology; p -factor; Pittsburgh Girls Study

(Received 31 July 2021; revised 30 March 2022; accepted 31 March 2022; First Published online 11 July 2022)

Introduction

The prevalence of comorbid/co-occurring psychopathology has presented conceptual and methodological challenges for studying mental illness. Prior conceptualizations that classify mental disorders as discrete categories are undermined by the staggering rates of co-occurring¹ psychopathology, with nearly 50% of those diagnosed with a mental disorder meeting criteria for at least one other disorder simultaneously (D. L. Newman et al., 1998). Because the co-variation of mental disorders has become the norm rather than the exception, the field has experienced a paradigm shift towards the use of broader transdiagnostic models, including a framework that theorizes disorders to stem from two latent vulnerabilities (i.e., internalizing and externalizing; Krueger et al., 1998).

The two-factor internalizing-externalizing structure has been extensively replicated, with these factors found to have moderate, positive correlations (e.g., M. D. Kramer et al., 2008; Krueger et al., 1998; Lahey et al., 2017). These observed correlations have in turn fostered researchers to search for a more global factor of

psychopathology – termed “ p ” (Caspi et al., 2014) – that may account for the phenotypic stability and co-occurrence of various mental disorders (Caspi & Moffitt, 2018). Although a number of studies have found evidence for a general p -factor using different symptoms and age groups (e.g., Brandes et al., 2019; Gomez et al., 2019; Greene & Eaton, 2017; Laceulle et al., 2015; Lahey et al., 2012, 2015; Tackett et al., 2013; Watts et al., 2019), extensive debate has persisted surrounding the interpretability and utility of p (see Smith et al., 2020 for a review).

Similar to the general factor of intelligence (g -factor; Spearman, 1904), dominant conceptualizations of p have shown preference towards causal, or common cause positions that theorize p as a latent vulnerability which influences one’s propensity for developing psychopathology (Aristodemou & Fried, 2020; Caspi et al., 2014; Caspi & Moffitt, 2018; Levin-Aspenson et al., 2021; van Bork et al., 2017). Support for common cause interpretations of p have largely been based on replications of the p -factor, as well as evidence suggesting p is moderately heritable and predictive of several adverse clinical outcomes (e.g., Conway et al., 2019; Lahey et al., 2017; M. M. Martel et al., 2017; Pettersson et al., 2018).

Nonetheless, proponents of p as a causal or substantive entity have not gone without criticism, and several concerns have been raised with respect to the methodology and “weak” theories used to justify the validity of p (Bonifay et al., 2017; Fried et al., 2021; Fried, 2020; van Bork et al., 2017; Watts et al., 2020). For example, given that structural models are not a rigorous test of causality, nor are intended to “discover” the existence of latent constructs (Bollen

¹For the purposes of this paper, co-occurrence rather than comorbidity is used as a more general term to describe the concurrent manifestation of two or more mental disorders (see Lilienfeld et al., 1994).

Corresponding author: Alexandria M. Choate, email: achoate@usf.edu

Cite this article: Choate, A. M., et al. (2023). The general psychopathology factor (p) from adolescence to adulthood: Exploring the developmental trajectories of p using a multi-method approach. *Development and Psychopathology* 35: 1775–1793, <https://doi.org/10.1017/S0954579422000463>



& Lennox, 1991; Borsboom et al., 2003), a number of papers have highlighted issues with reifying the p -factor as a causal entity based on model fit or findings of a strong general factor (Watts et al., 2020; van Bork et al., 2017). However, these critiques are not necessarily unique to the p -factor and could apply to any latent variable model.

An additional concern related to the p -factor stems from the quantity of studies arguing that p reflects a substantive construct without ruling out alternative explanations. Stated differently, a large proportion of studies have centered their research questions around the assumption that p is a valid construct that explains the manifestation of psychopathology, rather than a variable in need of explanation (Fried et al., 2021). Yet, whether the statistical emergence of p is produced by an unobserved vulnerability or is attributable to an entirely different data generating process (e.g., result of local symptom interactions, communal impairment, etc.), has remained unclear. Given the research and clinical implications that an agreed upon interpretation of p may offer (Levin-Aspenson et al., 2021), it will be important for future work to adequately test and rule out alternative explanations behind the general p -factor.

Although there are other alternative explanations of p that go beyond the scope of this paper (e.g., Oltmanns et al., 2018; van Bork et al., 2017), one intriguing hypothesis first introduced in the intelligence literature suggests that p is a product of evolving symptom interactions, rather than the cause of them (Caspi et al., 2014). This theory, termed dynamic mutualism, suggests that the positive manifold underlying the p -factor may be caused by developing interactions between lower-level symptoms and other psychological, biological, and environmental processes (van der Maas et al., 2006). These processes are assumed to be independent in early childhood, though are predicted to form increasingly strong associations throughout development until a state of equilibrium is reached (van der Maas et al., 2006, 2017). Accordingly, the core predictions of mutualism theory challenge common cause positions of p , in addition to other developmental psychopathology theories. For example, the differentiation hypothesis directly opposes predictions of dynamic mutualism by stipulating that disorder co-variation should decrease with age as symptoms become more differentiated from one another over time (Lahey et al., 2004; Lilienfeld et al., 1994; Sterba et al., 2010). This hypothesis has been extended to the p -factor under the term “ p -differentiation,” and is based on the assumption that p captures a general predisposition towards psychopathology that becomes increasingly specific with age (McElroy, Belsky, et al., 2018; Murray et al., 2016; Patalay et al., 2015).

To our knowledge, only two published studies have explicitly tested dynamic mutualism in the development of p by estimating variants of a bifactor model (McElroy, Belsky, et al., 2018; Murray et al., 2016). The bifactor model yields an indirect test of mutualism by evaluating whether the longitudinal strength and/or reliable variance accounted for by p increases, implying that the relations between lower-level symptoms are strengthening over time. Guided by this assumption, Murray et al. (2016) used an exploratory bifactor technique in a large, ethnically diverse sample of European children to extract a general p -factor and four specific factors based on teacher reports at eight time points (ages 7–15). Results were incongruent with mutualism, indicating that the strength of p and the specific factors were stable over time.

In a related study, dynamic mutualism theory was tested by estimating cross-sectional bifactor models across ages 2–14 using maternal self-reports on internalizing, externalizing, and attention-related symptoms. Results were in agreement with Murray

et al. (2016), revealing high factor stability and a dominant p -factor that accounted for the most variance across development (McElroy, Belsky, et al. 2018). The authors also assessed the phenotypic stability of p and the specific factors by saving factor scores from the bifactor models to later use in a cross-lagged panel design. Findings from this model suggested that cross-lags were attenuated and less consistent compared to the autoregressive paths, and that p both predicted, and was predicted by, the specific factors at various ages (McElroy, Belsky, et al. 2018).

Despite the value of these studies, several limitations still preclude strong inferences on the plausibility of mutualism theory in the development of p . First, previous studies have relied on teacher and/or parent reported data, which are often misaligned with self-report data from children or adolescents (T. L. Kramer et al., 2004; Van Roy et al., 2010). Additionally, the analyzed age ranges in recent studies focused on earlier periods of development, with less research examining transitional periods from adolescence to adulthood, or young adulthood to middle-adulthood. However, conceptualizing psychopathology from a mutualism perspective implies that certain developmental windows or “sensitive periods” may be characterized by unique symptom dynamics, such that the strength of these interactions or types of interactions may differ across development (Kievit, 2020). This suggests that tailoring interventions to a specific developmental window may be an effective tactic for intervention (Forbes et al., 2019). Equally, research that can identify which developmental periods or transitions are most vulnerable to the expression and/or escalation of symptoms may further inform preventative approaches to psychopathology.

Second, mutualism theory makes intraindividual level predictions that are suggested to inform phenomena at the between- and within-person levels (van der Maas et al., 2006, 2017). Previous investigations that have solely used cross-sectional approaches (e.g., bifactor or cross-sectionally derived panel models) are thus ill-equipped to examine mutualistic processes due to the conflation of within and between-person effects. Multilevel bifactor models address this concern by separating out these distinct variance components (Aitken et al., 2020; Constantinou et al., 2019), though assumptions tied to these models may be too strict or unrealistic to use with developmental data (e.g., measurement occasions are assumed to experience equivalent change).

Lastly, the types of cross-sectional models used in prior studies can pose other drawbacks for testing mutualistic processes if the data reflects both positive and negative interactions that can cancel out at the latent level. For instance, in the case of the cross-sectional bifactor model, changes in p may appear stagnant if some individuals display broader patterns of symptoms that become increasingly specific over time (e.g., higher levels of p lead to narrower symptom expressions), while others show a small range of symptoms that gradually expand throughout development (McElroy, Belsky, et al., 2018). Consequently, methods with developmentally appropriate assumptions that can adequately tease apart between- and within-person dynamics are needed to provide a more accurate investigation of mutualism theory and the p -factor.

The present study

The aim of the present study is to provide a rigorous and more theoretically appropriate test of dynamic mutualism theory by investigating its ability to explain the development of the p -factor in a racially diverse, population-based sample of girls who reported annually on various internalizing and externalizing pathologies

from ages 14–21.² Our decision to investigate mutualism theory during this developmental window is twofold: first, evidence strongly suggests that escalations in psychopathology tend to occur around early to mid-adolescence (Dalsgaard et al., 2020; Kessler et al., 2005), making this timeframe a reasonable starting point for examining mutualistic processes. This is especially true considering the evidence that suggests females have a higher propensity to develop psychopathology during adolescence compared to their male counterparts (Hayward, 2003; M. Martel, 2013; Ullsperger & Nikolas, 2017). Second, prior research has indicated that the transition from adolescence to young adulthood is one of the most sensitive periods of development due to the significant diversity in life paths, and widespread biological, psychological, and social role changes (e.g., Cicchetti & Rogosch, 2002; Schulenberg et al., 2004; Schulenberg & Zarrett, 2006; Zarrett & Eccles, 2006). The *p*-factor literature, however, has primarily focused on the development of psychopathology during childhood and/or early adolescence (Carragher et al., 2016; Deutz et al., 2020; Lahey et al., 2015; Olino et al., 2018; Patalay et al., 2015; Pettersson et al., 2018; Sallis et al., 2019; Snyder et al., 2017), neglecting important transitions that occur later in development.

In the present study, we tested predictions of dynamic mutualism theory by translating its core assumptions into three statistical models, including longitudinal bifactor models, random-intercept cross-lagged panel models (RI-CLPMs), and network models. While these statistical approaches are not intended to be directly compared, when taken together, results across models offer a detailed description of whether mutualistic processes are supported in the data. Evidence favoring dynamic mutualism theory was based on a series of statistical tests, in which we hypothesized that (1) the bifactor model would reveal a robust general *p*-factor that increased in strength and variance accounted for with age, (2) *p* and/or the internalizing-externalizing factors would have significant, mostly positive bidirectional effects at the between- and within-person levels,³ (3) including these bidirectional effects between *p* and the internalizing and externalizing factors in the bifactor models, or between internalizing and externalizing in the RI-CLPMs, would significantly improve model data fit, and (4) associations between internalizing and externalizing indices would strengthen with age, as evidenced by increasing estimates of centrality, edge weights, and small-worldness in the network models.

Method

Sample and procedure

Data were obtained from the Pittsburgh Girls Study ($N = 2450$), a prospective longitudinal study conducted in an urban setting in the greater Pittsburgh area (Keenan et al., 2010). Following an enumeration of the city of Pittsburgh, low-income neighborhoods were intentionally oversampled between 1999–2000 to increase the prevalence of girls' externalizing behavior. Out of the 2875 eligible

²Importantly, although Lahey et al. (2015) previously constructed a bifactor model to examine the longitudinal associations of *p* in the same population-based sample of girls, their study focused on a younger developmental period (ages 5–11) and utilized parental reports rather than self-report data. Therefore, the present study not only extends results of Lahey et al. (2015) but builds upon prior findings that have examined the development of the *p*-factor in younger ages with teacher or parent reports (McElroy, Belsky, et al., 2018; Murray et al., 2016).

³Of note, while positive, reciprocal interactions are predicted by dynamic mutualism, the presence of some negative or sparse interactions can still be consistent with mutualism and lead to a positive manifold in the data if a sufficient number of positive associations are present (van der Maas et al., 2017).

families, 85% agreed to study participation, resulting in a sample of 2450 girls divided between four age-specific cohorts (ages 5–8). Slightly over half of girls were African American/Black (52.92%), 41.17% were White, and 5.92% identified as a different race or multiracial (for more details on sampling procedures, see Hipwell et al., 2002).

The present analyses utilized eight consecutive annual waves of data (Waves 7–14) that spanned across ages 14–21. Data with sufficient cases were included in the analyses and resulted in a total sample of 2339 girls. At age 14, Wave 7 served as the initial data collection period for cohort 8 and Wave 10 served as the initial data collection period for cohort 5. The average sample retention rate across Waves 7–14 was high (87.26%), with the retention rate for each wave ranging from 86.25%–91.3%.

Attrition analyses, based on logistic regressions, suggested that families who did not receive public assistance or girls who identified as White at Wave 1 were significantly more likely to have incomplete data at age 21. Likewise, missingness at age 21 was significantly related to single-parent status, such that girls who were from a single-parent household at Wave 1 were more likely to remain in the study compared to girls who lived in a two-parent household. Other demographic variables at Wave 1, such as parental education and ethnicity, were not statistically different between those with and without missing data at age 21.

Procedure

Study procedures were approved by the University of Pittsburgh Human Research Protection Office. Caregivers provided written informed consent for study participation and girls provided assent until age 18, at which time girls then provided their own informed consent. Computerized assessments were completed separately by girls and their caregivers and families received monetary compensation for their participation (Hipwell et al., 2002).

Measures

Self-report measures capturing common internalizing and externalizing disorders and related constructs were used to assess for symptoms of attention-deficit hyperactivity disorder (ADHD), conduct disorder and antisocial personality disorder traits (CD/ASPD), generalized anxiety disorder (GAD), major depressive disorder (MDD), oppositional defiant disorder (ODD), and past year frequency of alcohol, marijuana, and tobacco use. Internal consistency was evaluated based on McDonald's (1999) coefficient Omega (ω), which serves as a more practical index of scale reliability compared to Cronbach's alpha (Dunn et al., 2014). Reliability across measures generally fell within an acceptable range for research purposes ($\omega > .70$; Hayes & Coutts, 2020) and can be found in Table 1.

The Adolescent Symptom Inventory-4 (ASI-4; Gadow & Sprafkin, 1999) and Adult Self-Report Inventory-4 (ASRI-4; Gadow et al., 2004) measured symptoms of ADHD, CD/ASPD traits, MDD, and ODD. CD traits were measured from ages 14–17 and were substituted with measures of ASPD traits from ages 18–21. The ASI-4 and ASRI-4 are rated based on past year symptoms using a 0–3 Likert scale with response choices: *never*, *sometimes*, *a lot*, and *all the time*.

The Nicotine, Alcohol and Drug Use scale (NADU; adapted from Pandina et al., 1984) was used to assess the frequency of past year alcohol, marijuana, and tobacco use (separately for each substance). The NADU is rated on an 8-point scale, where a "0" denotes *no past year use*, and a "7" signifies use of the substance

Table 1. Descriptive statistics and reliability

	ADHD	CD/ASPD	ODD	Substance Use	GAD	MDD
Age 14						
Mean	13.82	1.29	5.48	0.32	2.07	7.57
SD	7.99	2.10	4.06	0.86	1.51	4.88
ω	.90	.77	.87	.74	.78	.77
Age 15						
Mean	13.71	1.42	5.50	0.50	2.13	7.26
SD	8.10	2.12	4.04	1.08	1.60	4.86
ω	.91	.76	.87	.78	.80	.79
Age 16						
Mean	12.95	1.29	5.05	0.66	2.10	6.89
SD	7.97	2.07	3.87	1.22	1.60	4.84
ω	.91	.77	.87	.72	.81	.80
Age 17						
Mean	12.16	1.17	4.74	0.86	2.07	6.49
SD	8.13	1.94	3.91	1.37	1.66	4.93
ω	.91	.68	.87	.68	.82	.83
Age 18						
Mean	12.21	1.36	4.91	1.23	1.70	6.01
SD	7.99	1.73	3.93	1.54	1.19	4.44
ω	.91	.75	.88	.64	.81	.86
Age 19						
Mean	12.02	1.22	4.52	1.43	1.75	6.31
SD	7.90	1.66	3.77	1.60	1.21	4.60
ω	.91	.75	.88	.61	.81	.85
Age 20						
Mean	12.34	1.18	4.63	1.61	1.82	6.70
SD	8.18	1.69	3.88	1.64	1.25	4.73
ω	.92	.75	.88	.57	.82	.86
Age 21						
Mean	12.31	1.10	4.55	1.78	1.84	6.83
SD	7.96	1.51	3.68	1.61	1.25	4.64
ω	.91	.71	.87	.56	.82	.85

Note. ADHD = attention-deficit hyperactivity disorder; CD/ASPD = conduct disorder/antisocial personality disorder traits; GAD = generalized anxiety disorder; MDD = major depressive disorder; ODD = oppositional defiant disorder; Substance Use = average frequency of alcohol, marijuana, and tobacco use; SD = standard deviation; ω = total omega reliability coefficient.

every day or more than once a day. Alcohol use was defined as any consumption of beer, wine, or liquor. As a general proxy for past year substance use, a composite Substance Use score was computed at each age to represent the average frequency of alcohol, marijuana, and tobacco use. Reliability for this scale was generally lower compared to other scales, which may in part reflect the smaller number of scale items used to calculate reliability compared to other measures.

GAD was assessed using the Screener for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1997) from ages 14–17 and was replaced by the ASRI-4 (Gadow et al., 2004) and one item from the UCLA Loneliness Scale (ULS-20; D. Russell

et al., 1980; D. W. Russell, 1996) from ages 18–21. The SCARED is a self-report measure designed to screen for childhood anxiety disorders, such as GAD, separation anxiety disorder, panic disorder, and school phobia. Items are rated on a 3-point Likert scale consisting of the choices: *not true or hardly ever true*, *sometimes true*, and *very true*, respectively. Starting at age 18, girls were administered the anxiety module of the ASRI-4 instead of the SCARED screener. The ASRI-4 and the SCARED have considerable overlap, though the ASRI-4 focuses on generalized anxiety symptoms and consists of 15 questions that are rated on a 0–3 scale.

To ensure a fair comparison of anxiety symptoms across age, a subset of items that best represented GAD was selected from the SCARED and ASRI-4 measures, respectively. Items were matched based on wording and content, yielding a total of eight items from the SCARED, seven items from the ASRI-4 anxiety subscale, and one item from the ULS-20. The ULS-20 item, “is shy,” was specifically added as a parallel to the SCARED item, “I’m shy with people I don’t know well,” and was included with the other ASRI-4 items. Due to differences in Likert scaling between SCARED and ASRI-4 measures, retained anxiety items were re-scaled to be on the same metric using the proportion of maximum scaling method (Little, 2013). Reliability for the created GAD measures were within an acceptable range.

Data analytic plan

Dynamic mutualism theory was examined by translating its fundamental assumptions into three statistical models. In a recent update of the mutualism model, van der Maas et al. (2017) outlined a comprehensive network model of intelligence that incorporated four primary mechanisms to explain the development of cognitive ability: mutualistic coupling between lower-level cognitive processes, differences in centrality across cognitive processes (e.g., some processes may be more central than others, thereby influencing growth or development at a higher rate), sampling in cognitive test scores, and multiplier effects that are routed through the environment. The authors discuss the utility of network analysis in testing dynamic mutualism, though do not explicitly state that this analytical technique should be the only approach used to examine mutualism theory (van der Maas et al., 2017). As such, the estimated statistical models in the present study were selected based on prior developmental psychopathology and/or intelligence research that has demonstrated the utility of bifactor models, RI-CLPMs, and network models for surveying mutualistic processes (Hofman et al., 2018; Kan et al., 2019; Kievit et al., 2019; McElroy, Belsky, et al. 2018; Murray et al., 2016).

Measurement invariance

Longitudinal measurement invariance was assessed in the best fitting bifactor model and RI-CLPM to gauge whether the underlying meaning of the different internalizing and externalizing constructs was interpreted consistently across development (Widaman et al., 2010). More information about these procedures and results are presented in the supplemental materials (see Table S3).

Longitudinal bifactor models

Confirmatory bifactor models were estimated using the *lavaan* package in R-Studio (Rosseel, 2012) and included a general *p*-factor and two specific factors reflecting the internalizing and externalizing domains. Internalizing and externalizing composite scores (i.e., sum scores) for each measure served as a proxy for

construct/symptom severity and were fixed to load on either the internalizing or externalizing factor in addition to p . GAD and MDD were fixed to load on internalizing, while ADHD, CD/ASPD traits, frequency of substance use, and ODD were fixed to the externalizing factor. An average severity score was computed for CD/ASPD traits rather than a sum score to account for differences in the number of items used to measure CD and ASPD symptoms.

Bifactor models were estimated using robust maximum likelihood (MLR) to account for any data non-normality (Satorra & Bentler, 1994) and were identified by fixing factor variances to 1 with a mean of 0. Missing data due to sample attrition were handled with Full-Information Maximum Likelihood (FIML), which performs equally well, if not better, than other missing data techniques (e.g., multiple imputation; Larsen, 2011). Model fit was judged based on the AIC and robust variants of the CFI, TLI, and RMSEA that are corrected for non-normality (Brosseau-Liard & Savalei, 2014; Savalei, 2018). CFI and TLI $\geq .90$ and RMSEA $\leq .06$ were indicative of adequate model fit (Hu & Bentler, 1999; Little, 2013; Schermelleh-Engel et al., 2003). The chi-square goodness of fit statistic was reported, though was not of primary interest due its oversensitivity in larger samples (Floyd & Widaman, 1995). As a final check, we also estimated exploratory bifactor models to examine possible sources of misfit and cross-loadings (Marsh et al., 2014; Morin et al., 2020). Estimation procedures and model results are reported in the supplemental materials (see Tables S1 and S2).

Parallel to other research (e.g., Greene & Eaton, 2017; Olino et al., 2018; Snyder et al., 2017), we estimated two bifactor variants that provided information on between- and within-factor stability. Whereas the first model included only autoregressive paths for each factor (e.g., p at age 14 predicted itself at age 15), the second model introduced cross-lagged paths between p and the specific factors over time. In line with mutualism theory, we expected the bifactor with cross-lagged paths to demonstrate positive associations between the specific factors predicting p , but not vice versa.⁴ This model was in turn predicted to statistically outperform the bifactor with only autoregressive paths.

Factor strength, reliability, and replicability

Factor strength, model-based reliability, and construct replicability for p and the specific factors was quantified by calculating the explained common variance (ECV), Omega total (ω), Omega Subscale (ω_S), Omega Hierarchical/Hierarchical Subscale (ω_{HS}/ω_{HS}), Relative ω , and Hancock and Mueller's (2001) H construct replicability index (Reise, 2012; Rodriguez et al., 2016a, 2016b). All indices were calculated using the Microsoft Excel Bifactor Indices Calculator (Dueber, 2017) and were derived from the confirmatory and exploratory bifactor models at each age.

The ECV provides an index of factor strength and is the proportion of common variance explained by a given factor (Reise et al., 2013; Rodriguez et al., 2016a, 2016b; Sijtsma, 2009; Stucky & Edelen, 2014). ω is an index of model-implied reliability that returns the proportion of common variance across all factors relative to the total variance. ω_S is related to ω , though returns the proportion of variance in observed subscale scores that is explained by the general factor and a given specific factor. Of particular interest, ω_H is the percentage of systematic variance in raw total scores that is attributable to the general factor after controlling for the

influence of the specific factors. ω_{HS} is the specific factor version of ω_H and reflects the percentage of variance attributable to a specific factor after accounting for the variance explained by the general factor (McDonald, 1999; Reise et al., 2013; Rodriguez et al., 2016b; Zinbarg et al., 2005). Relative ω reflects the proportion of reliable variance in total (or subscale) scores that are attributable to the general factor or a specific factor, respectively (Rodriguez et al., 2016a). Lastly, the construct replicability index (H) assesses the replicability of the modeled factors by evaluating how well a latent variable is defined by its indicators. Values of $H \geq .70$ indicate that the factor is well-represented by its respective items and is likely to be replicable by other studies (Hancock & Mueller, 2001; Rodriguez et al., 2016a, 2016b).

To be consistent with mutualism theory, estimates of ω_H and ECV for the p -factor are predicted to progressively increase over time, with p capturing substantially more variance compared to the specific factors. Support against mutualism, in contrast, is evidenced by stagnant estimates of strength and/or variance accounted for by the p -factor over time. In determining whether the strength and/or variance explained by p meaningfully increased, Wald tests were used to statistically evaluate if changes in ω_H or ECV were significant. In other words, because both ω_H and ECV can be calculated from the estimated factor loadings, we constrained loadings on p to be equal at the first and last time point (i.e., ages 14 and 21) and used Wald tests to assess for any significant differences. If Wald tests were significant, factor loadings at age 14 and 21 were inspected to determine the direction of this difference.

Random-intercept cross-lagged panel models (RI-CLPMs)

The RI-CLPM allowed the relationships between internalizing and externalizing to be examined at the within-person level, independent of the p -factor. The RI-CLPM can be thought of as an extension of the traditional cross-lagged panel model, except most parameters are interpreted at the within-person level by including a random-intercept factor (Berry & Willoughby, 2017; Hygen et al., 2020). In other words, after accounting for the more stable between-person differences captured in the random-intercepts, autoregressive paths reflect the extent that deviations above or below one's personal average (i.e., expected score) carry-over into the next measurement occasion. Similarly, cross-lagged paths quantify whether person-specific deviations in one domain predict comparable deviations in a separate domain, while within-time factor correlations capture the degree that person-specific deviations at the same measurement period are related between domains (Berry & Willoughby, 2017; Hamaker et al., 2015).

RI-CLPMs were constructed in the *lavaan* package (Rosseel, 2012) with MLR and FIML estimation and were evaluated using the same goodness of fit criteria as the bifactor models. However, the RI-CLPMs did not include a general p -factor, as the purpose of this model was to determine if interactive effects were present between the internalizing and externalizing factors that would otherwise be subsumed by p . The first model served as a baseline model where autoregressive paths for the internalizing and externalizing factors were freely estimated but cross-lagged paths were constrained to zero. This model was expected to provide the poorest fit to the data if predictions of mutualism are supported. Next, we estimated two unidirectional models (i.e., internalizing predicting change in externalizing or vice versa), which was followed by the bidirectional model that served as a proxy for mutualism. Chi-square difference tests and the AIC were used to compare the nested RI-CLPMs, and parameters from the

⁴In the event that temporal associations were found to be driven by p predicting the specific factors, this was interpreted as evidence against mutualism.

superior fitting model were further inspected. Support for dynamic mutualism theory was based on whether the bidirectional RI-CLPM outperformed all other models, with the internalizing and externalizing factors expected to have mostly positive, significant cross-lagged associations over time. Insignificant and/or predominantly negative associations between domains were interpreted as evidence against mutualism theory.

Network models

As a final probe of dynamic mutualism, we used network analysis to examine reciprocal associations between internalizing and externalizing constructs independent from their latent domains. In doing so, we estimated weighted, unregularized networks as Gaussian Graphical Models (GGM; Lauritzen, 1996) using the *bootnet* package (i.e., “ggmModSelect”) in R-studio (Epskamp, 2015). GGMs were purposely not regularized, as unregularized estimation procedures are shown to outperform regularized networks when sample sizes are large and a small set of nodes are estimated (Foygel & Drton, 2010; Friedman et al., 2008; Williams & Rast, 2020; Williams et al., 2019). Under GGM estimation, a graphical LASSO algorithm is implemented that iteratively re-estimates the network without regularization using maximum likelihood estimation. This algorithm yields the most parsimonious model by adding or removing edges until the extended Bayesian Information Criteria no longer improves (Isvoranu & Epskamp, 2021).

To mirror the separation of processes in the RI-CLPMs, networks were estimated at the between- and within-person levels and missing data was handled with FIML. Whereas between-person networks explain the covariance patterns of stationary means *across* individuals, within-person networks detail the covariances of stationary means *within* individuals (Epskamp, Waldorp, et al., 2018). Within-subject networks were constructed using an approach outlined by Costantini et al. (2019), where each subject's grand mean is computed per node and subtracted from the subject's observed score at a given age.

Akin to other developmental work (e.g., McElroy, Shevlin, et al., 2018), we estimated three cross-sectional networks that were equidistant in time. This resulted in two networks (i.e., one between and one within-person) estimated at ages 14, 17, and 20. Age 20 was selected as the final estimated age rather than age 21 to ensure network comparisons were equally spaced. Due to the computational complexity required to estimate a symptom-level network, nodes were based on the same composite scores used in the bifactor models and RI-CLPMs. Average substance use frequency was the exception, and each substance was modeled as its own node.⁵

The *bootnet* package was used to assess the accuracy of the edge weights by calculating 95% confidence intervals (CIs) around the edges with non-parametric bootstrapping (Epskamp, 2015). Centrality stability was subsequently inspected via a case-drop bootstrapping approach, which allowed the correlation stability coefficient (CS-coefficient) to be subsequently estimated. The CS-coefficient reflects the proportion of cases that can be dropped to maintain a correlation of at least .70 between the original and bootstrapped network. CS-coefficients above .70, .50, and .25 indicate excellent, good, and fair stability, respectively (Epskamp, Borsboom, et al., 2018). In addition, bootstrapped difference tests

⁵Compared to the latent variable models, the reduced computational complexity in estimating the network models allowed for the average frequency of substance use to be separated into past year frequency of alcohol, marijuana, and tobacco use. Doing so enabled the associations across specific substances, as well as their relations with other internalizing and externalizing nodes, to be directly assessed in the network models.

were conducted for centrality metrics and edges to gauge whether these indices significantly differed from one another. The difference test results can be found in the supplemental materials (Figures S3–S6).

Node importance was judged by estimating centrality measures of closeness, betweenness, and expected influence (EI). Closeness refers to the average shortest path length between different pairs of nodes and quantifies the *indirect* influence of a given node. Betweenness is the number of times a node falls on the shortest path between two other nodes. Thus, nodes with high betweenness are often interpreted as bridges that foster connections between other nodes (Costantini et al., 2015, 2019; Newman, 2010; Opsahl et al., 2010). EI is similar to the strength centrality metric and supplies an index of how influential a node is in the entire network structure. Unlike strength, however, EI takes into account both positive and negative edge weights in its calculation (McNally, 2016; Robinaugh et al., 2016).

The small-worldness index (SWI) was obtained using the *qgraph* package (Epskamp et al., 2012) and is computed based on the average shortest path length and overall transitivity of the network (Newman, 2010). Networks characterized by high degrees of small-world properties have SWI values greater than 1 (with stricter cutoffs of 3 or more; Humphries & Gurney, 2008) and are sensitive to fluctuations in the network, such that changes in a single node are more likely to influence other nodes in the network (Borsboom et al., 2011). If predictions of mutualism are supported, the SWI is expected to exceed 1 and gradually increase throughout development. If the SWI is instead found to be small or weaken over time, this suggests that the network structure is becoming sparser with age and is inconsistent with mutualism theory.

Network comparisons

The *NetworkComparisonTest* package in R-studio was used to test whether networks differed in structure or global connectivity (van Borkulo et al., 2016). Structural invariance is reflected in the *M* test statistic and is the maximum difference across edges between two networks. Global invariance compares differences in the overall strength (i.e., the absolute value of the sum of edges) of two networks and is reflected in the *S* test statistic (Opsahl et al., 2010; van Borkulo et al., 2017). If significant differences at the structural or global level were observed, edge weight difference tests with a holm *p*-value adjustment were used to determine which edges statistically differed between networks (van Borkulo et al., 2017). Although NCTs can be used with dependent samples, the algorithm for dependent samples is still undergoing validation (van Borkulo et al., 2016). Therefore, NCTs were supplemented by correlating edges and centrality metrics across networks to further gauge differences in the network structures.

Results

Means, standard deviations (SDs), and reliability for each of the internalizing and externalizing measures are presented in Table 1. In brief, the best fitting bifactor and RI-CLPM were found to be partially invariant at the scalar level, with the bifactor demonstrating greater variability in factor loadings compared to the RI-CLPM. Globally, violations of measurement invariance appeared to be small and were mostly attributable to fluctuations in CD/ASPD traits and mean-level changes in the average frequency of substance use over time. ODD similarly exhibited some degree of metric non-invariance in the bifactor model, with its

loadings on p decreasing over time. Additional information on measurement invariance procedures and results can be found in the supplemental materials (Table S3).

Longitudinal bifactor models

Exploratory models

Standardized factor loadings from the exploratory models are reported in the supplemental materials (Table S1). Cross-loadings for internalizing and externalizing were generally small and fell below .20 across age. Estimates of factor strength, reliability, and replicability are similarly presented in the supplementals (Table S2). Estimates of factor strength and reliability for the p -factor were similar between confirmatory and exploratory models; however, exploratory models suggested that the variance accounted for by the specific factors was substantially weaker relative to the confirmatory models (described below).

Strength, reliability, and construct replicability of p and the specific factors

Factor strength, reliability, and construct replicability based on confirmatory bifactor models at each age are reported in Table 2. ECV and ω_H suggested that the strength and proportion of variance explained by p steadily increased throughout development, reaching its peak value at age 21 (ECV = .74; ω_H = .76). Relative ω for p similarly reached its peak value at age 21 (relative ω = .86), suggesting that 86% of the reliable variance in total scores can be attributed to the p -factor (Rodriguez et al., 2016b). Construct replicability for the p -factor was also high, with H above recommends cutoffs of $\geq .70$ at all ages. After controlling for the influence of the p -factor, ω_{HS} indicated that the variance attributable to the specific factors was substantially weaker. Whereas internalizing accounted for a greater proportion of variance at each age relative to externalizing (ω_{HS} = .21–.35), internalizing steadily decreased in the amount of variance it explained over time while externalizing remained stable (ω_{HS} = .18–.22).

In determining whether these increases in strength and/or reliable variance for the p -factor were statistically meaningful, equality constraints were imposed on each respective factor loading on p at the first and last time point (i.e., age 14 and age 21). Significant Wald tests were documented for ADHD (Wald test $\Sigma^2(1) = 49.78$, $p < .001$), CD/ASPD traits (Wald test $\Sigma^2(1) = 35.65$, $p < .001$), MDD (Wald test $\Sigma^2(1) = 37.40$, $p < .001$), and ODD (Wald test $\Sigma^2(1) = 149.67$, $p < .001$). In contrast, GAD (Wald test $\Sigma^2(1) = 0.91$, $p = .34$) and substance use frequency (Wald test $\Sigma^2(1) = 1.20$, $p = .27$) resulted in non-significant differences between ages. Inspection of factor loadings suggested that ADHD, CD/ASPD, and MDD displayed stronger loadings on p over time, while loadings for ODD marginally decreased by age 21.

Model fit and factor stability

Factor loadings and standard errors for the longitudinal bifactor models are presented in the supplemental materials (Table S4). The bifactor model with only autoregressive paths provided excellent fit to the data (SB- $\chi^2(df) = 1939.79(843)$, $p < .001$; R-CFI = .98; R-TLI = .97; R-RMSEA [90% CI] = .026 [.025–.028]; AIC = 322,793), with most loadings on the p -factor significant over time ($p < .001$). We next compared this model to a similar bifactor structure that included cross-lagged paths between p and the specific factors. This bifactor variation with cross-lagged paths had excellent fit to the data (SB- $\chi^2(df) = 1633.08(801)$, $p < .001$; R-CFI = .98; R-TLI = .98; R-RMSEA

Table 2. Factor strength, reliability, and replicability based on confirmatory bifactor models at each age

Age	Factor	ECV	ω_H/ω_{HS}	Relative ω	ω/ω_S	H
14	P -Factor	.63	.69	.82	.84	.84
	Internalizing	.52	.32	.54	.59	.37
	Externalizing	.31	.18	.21	.84	.66
15	P -Factor	.59	.66	.78	.84	.83
	Internalizing	.55	.35	.58	.61	.40
	Externalizing	.36	.21	.25	.84	.73
16	P -Factor	.61	.66	.79	.83	.82
	Internalizing	.53	.32	.54	.60	.37
	Externalizing	.34	.21	.26	.83	.62
17	P -Factor	.65	.68	.81	.84	.84
	Internalizing	.42	.27	.42	.63	.33
	Externalizing	.33	.21	.26	.82	.62
18	P -Factor	.69	.72	.82	.88	.90
	Internalizing	.34	.27	.33	.80	.39
	Externalizing	.29	.22	.27	.82	.50
19	P -Factor	.71	.73	.83	.88	.89
	Internalizing	.33	.26	.32	.81	.39
	Externalizing	.27	.21	.26	.82	.47
20	P -Factor	.72	.75	.84	.89	.89
	Internalizing	.31	.25	.31	.82	.37
	Externalizing	.27	.20	.24	.83	.47
21	P -Factor	.74	.76	.86	.88	.88
	Internalizing	.26	.21	.26	.82	.31
	Externalizing	.26	.19	.23	.82	.45

Note. ECV = Explained Common Variance; ω_H/ω_{HS} = Omega Hierarchical and Subscale Omega Hierarchical; ω/ω_S = Omega and Omega Specific; Relative ω = relative Omega; H = construct replicability.

[90% CI] = .023 [.022–.025]; AIC = 322,444), with the inclusion of these paths significantly improving model fit ($\Delta\chi^2(\Delta df) = 241.21(42)$, $p < .001$).

Estimates of within- and between-factor stability (i.e., autoregressions and cross-lags, respectively) can be found in Table 3. When cross-lagged effects were included, the p -factor was determined to have weaker temporal stability at later ages, with strong stability throughout adolescence ($\beta = .29-.72$, $p < .01$). Autoregressive paths for externalizing were larger, on average, compared to both p and internalizing, with these effects generally increasing with age ($\beta = .48-.80$, $p < .001$). Conversely, temporal stability for internalizing was more variable and tended to decline with age ($\beta = .51-.86$, $p < .05$). Autoregressive paths were mostly significant, apart from internalizing at age 17 predicting internalizing at age 18 ($\beta = .88$, $p = .11$).

In line with dynamic mutualism theory, the specific factors were found to significantly predict p at several ages, with within-factor stability (i.e., autoregressions) for p declining once cross-lagged paths were specified. Cross-lags were especially pronounced for internalizing predicting p , such that internalizing significantly predicted p at each age with these effects increasing over time ($\beta = .10-.46$, $p < .05$). In contrast, externalizing significantly predicted p at age 18 ($\beta = .26$, $p < .01$) and age 19 ($\beta = .42$, $p < .01$),

Table 3. Autoregressive and cross-lagged paths for the longitudinal bifactor model

	Est (SE)	β	Est (SE)	β	Est (SE)	β
<i>Age 14 → 15</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	1.05 (0.06)	.72***	0.22 (0.12)	.13	0.22 (0.13)	.16
Internalizing	0.14 (0.06)	.10*	1.30 (0.17)	.76***	-0.20 (0.06)	-.14**
Externalizing	0.16 (0.09)	.11	-0.39 (0.09)	-.23***	0.98 (0.15)	.68***
<i>Age 15 → 16</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.79 (0.06)	.72***	0.13 (0.11)	.09	0.05 (0.11)	.05
Internalizing	0.12 (0.05)	.12*	1.07 (0.19)	.86***	-0.17 (0.05)	-.19***
Externalizing	0.18 (0.10)	.16	0.01 (0.09)	.004	0.66 (0.12)	.65***
<i>Age 16 → 17</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.55 (0.06)	.63***	0.06 (0.10)	.04	0.54 (0.09)	.51***
Internalizing	0.14 (0.06)	.21*	0.88 (0.21)	.86***	-0.22 (0.09)	-.27*
Externalizing	-0.02 (0.07)	-.02	-0.11 (0.08)	-.08	0.56 (0.09)	.48***
<i>Age 17 → 18</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.38 (0.08)	.38***	0.35 (0.23)	.15	0.38 (0.12)	.21**
Internalizing	0.29 (0.09)	.45***	1.34 (0.83)	.88	-0.21 (0.13)	-.18
Externalizing	0.22 (0.07)	.26**	-0.13 (0.23)	-.07	1.22 (0.26)	.80***
<i>Age 18 → 19</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.30 (0.11)	.29**	0.01 (0.09)	.01	0.31 (0.09)	.23***
Internalizing	0.20 (0.09)	.46*	0.37 (0.18)	.76*	-0.03 (0.06)	-.05
Externalizing	0.24 (0.08)	.42**	-0.02 (0.07)	-.03	0.56 (0.11)	.76***
<i>Age 19 → 20</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.44 (0.10)	.44***	0.21 (0.13)	.21	0.24 (0.13)	.16
Internalizing	0.30 (0.10)	.33**	0.61 (0.13)	.65***	-0.33 (0.14)	-.24*
Externalizing	0.19 (0.10)	.24	-0.02 (0.12)	-.03	0.92 (0.20)	.78***
<i>Age 20 → 21</i>	<i>P-factor</i>		<i>Internalizing</i>		<i>Externalizing</i>	
P-factor	0.51 (0.12)	.46***	0.41 (0.16)	.38**	0.31 (0.15)	.20*
Internalizing	0.44 (0.16)	.42**	0.53 (0.19)	.51**	-0.20 (0.20)	-.14
Externalizing	0.13 (0.10)	.19	-0.14 (0.09)	-.20	0.79 (0.21)	.80***

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

and was itself predicted by p at ages 17–19 ($\beta = .21-.42$, $p < .05$) and age 21 ($\beta = .20$, $p < .05$). Significant cross-lags were further documented for internalizing negatively predicting future levels of externalizing between ages 15–17 ($\beta = -.27$ to $-.14$, $p < .05$) and again at age 20 ($\beta = -.24$, $p < .05$), though externalizing in turn only predicted internalizing at age 15 ($\beta = -.23$, $p < .001$).

Random-intercept cross-lagged panel models

Model fit and comparisons

Fit statistics for the RI-CLPMs are presented in Table 4 and suggested that all model variations demonstrated excellent fit to the data (CFI and TLI $> .95$, RMSEA $< .08$). Although R-CFI, R-TLI, and R-RMSEA produced equivalent estimates across models, chi-square difference tests and the AIC suggested that the mutualism model with bidirectional effects provided the best fit to the data. Parameter estimates for this model are reported in Table 5.

Interpretation of parameter estimates from the mutualism model

Parameters from the mutualism RI-CLPM indicated significant between-person variability as evidenced by the random-intercept factor variances. Stated differently, variances of the random-intercepts were found to statistically differ from 0, indicating significant between-person variability for both internalizing ($\sigma^2 = 5.23$, $p < .001$) and externalizing ($\sigma^2 = 10.07$, $p < .001$) domains. These between-person components were also significantly correlated (σ^2_B : $r = .78$, $p < .001$), implying that girls who scored above average on internalizing were more likely to score above average on externalizing throughout development.

At the within-person level, significant autoregressive paths were documented for internalizing across age ($\beta = .29-.58$, $p < .05$), providing some evidence for within-person carry-over effects. In other words, girls who deviated from their expected score in internalizing at a given age (i.e., girls who scored either above or below their personal average) were more likely to show similar deviations in internalizing at subsequent ages. Significant autoregressive effects also

Table 4. Goodness of fit and model comparisons for the random-intercept cross-lagged panel models (RI-CLPMs)

Model	SB- χ^2 (YB)	df	R-CFI	R-TLI	R-RMSEA	AIC
1. Baseline (INT \neq EXT)	2684.91*** (1.28)	887	.96	.95	.033	323698
2. Unidirectional (INT \rightarrow EXT)	2648.73*** (1.28)	880	.96	.95	.033	323655
3. Unidirectional (INT \rightarrow EXT)	2653.65*** (1.28)	880	.96	.95	.033	323672
4. Mutualism (INT \leftrightarrow EXT)	2630.18*** (1.28)	873	.96	.95	.033	32364
Model comparisons	$\Delta\chi^2$	Δdf	ΔR -CFI	ΔR -TLI	ΔR -RMSEA	Δ AIC
1 versus 2	32.14***	7	0	0	0	0
1 versus 3	31.34***	7	0	0	0	0
1 versus 4	52.97***	14	0	0	0	0
2 versus 4	18.45*	7	0	0	0	0
3 versus 4	22.88**	7	0	0	0	0

Note. INT = internalizing factor; EXT = externalizing factor; df = degrees of freedom; SB- χ^2 = Satorra-Bentler corrected chi-square statistic; YB = Yuan-Bentler correction; R-CFI = robust comparative fit index; R-TLI = Robust Tucker-Lewis index; R-RMSEA = robust root-mean-square error of approximation; AIC = Akaike Information Criterion; $\Delta\chi^2$ = change in chi-square based on non-robust chi-square statistic.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 5. Parameter estimates for the bidirectional random-intercept cross-lagged panel model (RI-CLPM)

	Age 15		Age 16		Age 17		Age 18		Age 19		Age 20		Age 21	
	Est (SE)	β	Est (SE)	β	Est (SE)	β	Est (SE)	β	Est (SE)	β	Est (SE)	β	Est (SE)	β
<i>Autoregressions</i>														
INT \rightarrow INT	0.48 (0.11)	.56***	0.53 (0.10)	.58***	0.46 (0.12)	.43***	0.66 (0.15)	.58***	0.30 (0.13)	.29*	0.43 (0.14)	.40**	0.50 (0.21)	.54*
EXT \rightarrow EXT	0.52 (0.06)	.57***	0.55 (0.06)	.60***	0.40 (0.06)	.43***	0.19 (0.10)	.20*	0.23 (0.11)	.25*	0.16 (0.15)	.15	-0.18 (0.25)	-.17
<i>Cross-lags</i>														
INT \rightarrow EXT	-0.01 (0.05)	-.02	-0.09 (0.07)	-.10	-0.03 (0.08)	-.03	0.06 (0.08)	.07	0.03 (0.07)	.04	0.18 (0.09)	.27*	0.48 (0.17)	.73***
EXT \rightarrow INT	-0.21 (0.08)	-.23**	-0.14 (0.06)	-.16*	-0.10 (0.08)	-.09	-0.37 (0.16)	-.28*	0.06 (0.17)	.04	0.13 (0.22)	.07	0.10 (0.31)	.07
<i>Correlations</i>							Est (SE)	<i>r</i>						
Between-person							5.67 (0.29)	.78***						
Age 14							4.56 (0.38)	.60***						
Age 15							2.77 (0.29)	.63***						
Age 16							2.28 (0.26)	.65***						
Age 17							2.28 (0.29)	.80***						
Age 18							3.75 (0.28)	.82***						
Age 19							3.75 (0.25)	.80***						
Age 20							4.15 (0.31)	.85***						
Age 21							3.43 (0.33)	.93***						

Note. Est = unstandardized beta; SE = standard error; β = standardized beta; INT = Internalizing; EXT = Externalizing.
* $p < .05$; ** $p < .01$; *** $p < .001$.

emerged for externalizing but were less stable compared to internalizing and were only significant from ages 15–19 ($\beta = .20-.60$, $p < .05$). Thus, compared to the autoregressions in the bifactor model that indicated externalizing to be the most stable factor over time, within-person estimates found the opposite pattern of results.

In contrast, cross-lagged effects were mostly non-significant for internalizing predicting change in externalizing throughout

adolescence, implying that person-specific deviations in externalizing symptoms were not dependent upon prior deviations in internalizing. This pattern, however, did shift by adulthood, such that internalizing significantly predicted within-person change in externalizing at ages 20 ($\beta = .27$, $p < .05$) and 21 ($\beta = .73$, $p < .01$). Relatedly, cross-lagged paths from externalizing predicting internalizing were negative during adolescence but became

positive, albeit non-significant, by early adulthood. Specifically, externalizing significantly predicted within-person change in internalizing at age 15 ($\beta = -.23, p < .01$), age 16 ($\beta = -.16, p < .05$), and age 18 ($\beta = -.28, p < .05$). This indicated that within-person deviations in internalizing, at least during these ages, can be predicted by an individual's prior deviation from their expected score in externalizing (e.g., girls who reported above average symptoms of externalizing, relative to their personal average, were more likely to report fewer internalizing symptoms a year later). Within-time factor correlations were also positive and linearly increased with age, implying that girls who scored above or below their personal average on internalizing showed comparable deviations in externalizing at the same measurement occasion.

As an added check, we also examined the extent that effects in the mutualism RI-CLPM changed when between-person variability was not controlled for by constraining the variances and covariances of the random-intercept factors to 0. This resulted in a nested model under the RI-CLPM that is equivalent to the cross-lagged panel model (CLPM; Hamaker et al., 2015). Constraining the random-intercept variances and covariances resulted in significantly poorer fit to the data based on the chi-square test⁶ ($P(x^2 = 6.46, p < .001)$), which further supported that these domains were characterized by meaningful differences across individuals (Stoel et al., 2006). Intriguingly, when these between-person effects were not directly accounted for, autoregressive paths became more pronounced for internalizing ($\beta = .77-.89, p < .001$) and externalizing ($\beta = .63-.78, p < .001$). Cross-lagged effects from externalizing to internalizing also decreased in frequency, such that externalizing only predicted internalizing at age 15 ($\beta = -.16, p < .01$) but was positively predicted by internalizing at age 15 ($\beta = .09, p < .05$), age 17 ($\beta = .11, p < .01$), and age 18 ($\beta = .12, p < .05$).

Network models

Graphs of the between- and within-person networks for ages 14, 17, and 20 can be found in Figure 1. For ease of visual comparison, networks are presented using the Fruchterman and Reingold (1991) graphing algorithm (i.e., the "spring" layout in the *qgraph* package) using the same average layout (Epskamp et al., 2012). More densely connected nodes, which are represented as circles, are concentrated together towards the middle of the graph.

Accuracy and stability of networks

Nodes with stronger edges were found to have smaller CIs that did not overlap with zero, implying that these edges were more accurate compared to weaker edges that generally had larger CIs (Epskamp, Borsboom, et al., 2018). For example, node pairs such as ADHD-ODD, GAD-MDD, MDD-ADHD, and frequency of alcohol use with other substance use nodes generally had smaller CIs that did not overlap with zero. In contrast, node pairs such as MDD-CD/ASPD tended to have weaker edges with wider CIs (see Figure S1).

Centrality stability for the between-person networks was fair for betweenness (CS-coefficients > 0.50), with good metric stability found for closeness and EI (CS-coefficients > 0.70). Within-person networks suggested that stability was poor for betweenness

(CS-coefficients < 0.25), though ranged from adequate to good for closeness (CS-coefficients $> 0.50-0.70$) and EI (CS-coefficients > 0.70). Estimates of betweenness were thus not interpreted for within-person networks, and node centrality was evaluated based on closeness and EI. More information on the accuracy and stability of the networks are reported in the supplemental materials (Figures S1–S6).

Between-person networks

Nodes associated with the internalizing or externalizing domains generally clustered together, with substance use nodes forming their own cluster separate from the externalizing nodes. Several positive edges were documented between the various internalizing and externalizing nodes, which was consistent with small-worldness estimates above 1 – but not above stricter cutoffs of 3 ($SWI_{14} = 1.07$; $SWI_{17} = 1.27$; $SWI_{20} = 1.47$).

Standardized centrality estimates by age are displayed in Figure 2. Correlations between centrality indices ranged from moderate to large and can be found in the supplemental materials (Table S5). ADHD, CD/ASPD, and ODD tended to be more central in the network, while GAD and substance use related nodes were least central overall (Figure 2). ADHD was determined to be the most central node overall due to its positive associations with other internalizing and externalizing nodes, high estimates across the different centrality measures, and its centralized position in the network structure (Figure 1).

Edge weights were found to generally strengthen over time apart from ODD and substance use nodes. ADHD had the largest number of positive edges and was significantly related to CD/ASPD, ODD, GAD, and MDD across age. Pearson correlations for the estimated edges were largest between ages 14 and 17 ($r = .94$), followed by ages 17 and 20 ($r = .85$), and ages 14 and 20 ($r = .79$). Structural differences and edge weight differences were non-significant between ages 14 and 17 ($M = 0.12, p = .30$), though were significant between ages 14 and 20 ($M = 0.19, p < .001$), and ages 17 and 20 ($M = 0.19, p < .01$). Results of the edge weight difference tests indicated that MDD-GAD, MDD-ADHD, and ADHD-CD/ASPD edges significantly increased with age, while ODD-ADHD, ODD-CD/ASPD traits, marijuana and tobacco use, and alcohol and tobacco use edges decreased over time (see Table S7 in the supplemental materials). Notwithstanding the reported structural non-invariance and increases in small-worldness, the global connectivity of the network (i.e., sum of all edges) remained stable over time. That is to say that while the broader levels of psychopathology were consistent, the structural patterns and manifestations of these pathologies appeared to fluctuate to some degree.

Within-person networks

Networks at the within-person level had comparable clustering patterns to the between-person structures (see Figure 1). Further, although the SWI of the within-person networks was characterized by sharper increases, both between- and within-person structures reached similar levels of small-worldness by age 20 ($SWI_{14} = 1.16$; $SWI_{17} = 1.47$; $SWI_{20} = 1.44$). This suggested that the density of the network increased throughout development, such that the activation of one node was more likely to have downstream effects on other nodes in the network (Borsboom et al., 2011).

Centrality estimates also mirrored the between-person networks, such that ADHD, ODD, and CD/ASPD tended to be the most influential in the network, followed by MDD at later ages

⁶The chi-square test was used over the chi-square difference test, as this is a more appropriate test of differences in model fit when constraints are imposed on the bound of a given parameter (Stoel et al., 2006).

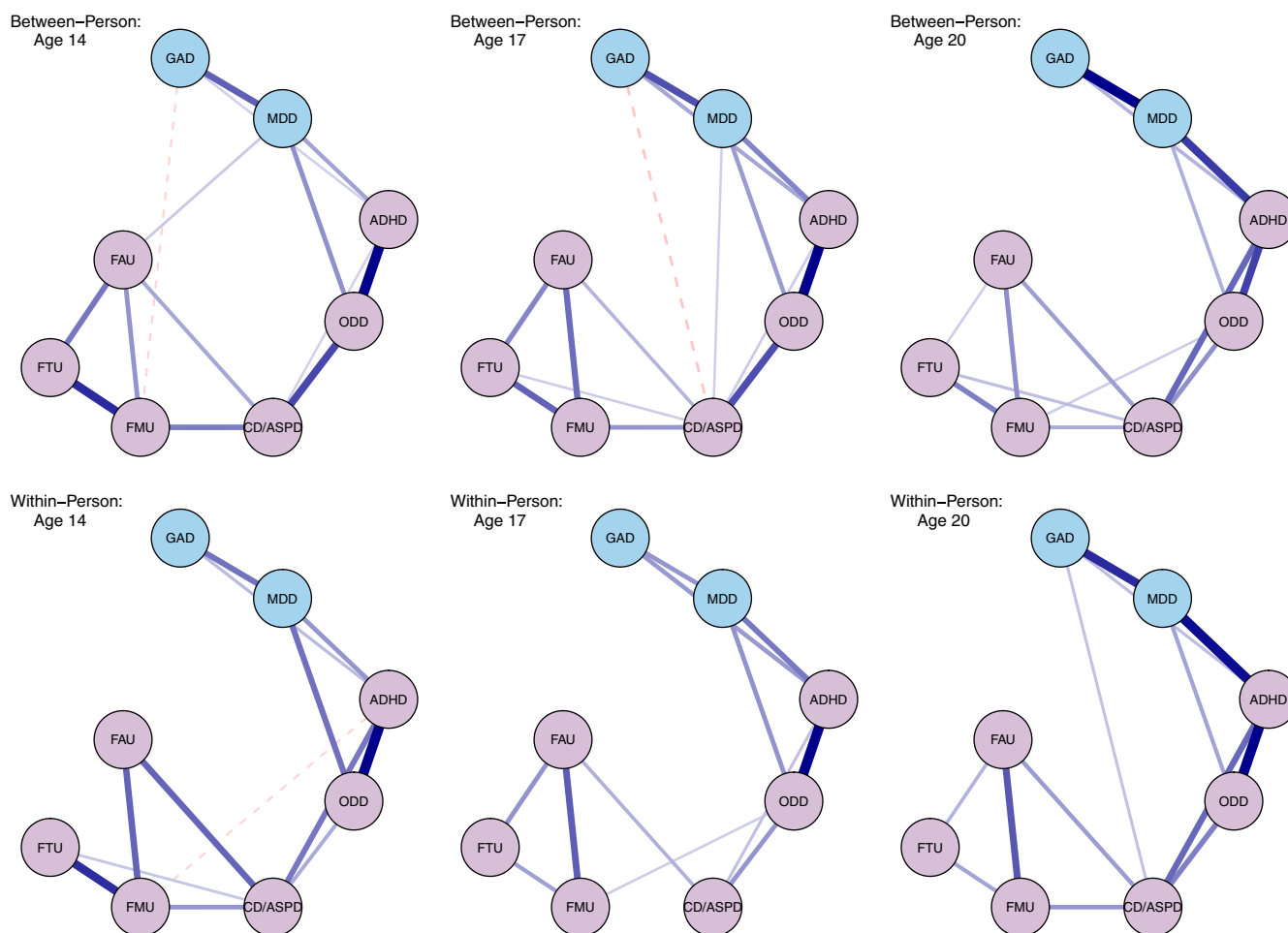


Figure 1. Between- and within-person network graphs by age. Nodes and edges are represented by circles and lines, respectively. Thicker lines indicate stronger associations between two nodes after controlling for all other associations in the network. ADHD = attention-deficit hyperactivity disorder; CD/ASPD = conduct disorder/antisocial personality disorder traits; FAU = frequency of alcohol use; FMU = frequency of marijuana use; FTU = frequency of tobacco use; GAD = generalized anxiety disorder; MDD = major depressive disorder; ODD = oppositional defiant disorder.

(Figure 2). Correlations amongst centrality measures, on average, were greatest for ages 14 and 20, with weaker correlations between ages 14 and 17 (see supplemental Table S6). Analogous to the between-person networks, positive edges were found between GAD-ADHD, MDD-ADHD, and MDD-ODD at all ages, implying that ADHD may serve as a bridge node between other internalizing and externalizing nodes.

Edges were also highly correlated across age (Ages 14 and 17: $r = .82$; Ages 17 and 20: $r = .87$; Ages 14 and 20: $r = .84$), though exhibited greater oscillations relative to the between-person networks. For instance, despite most edges in the between-person networks progressively increasing with age, several edges in the within-person structures decreased in strength from age 14 to age 17, though increased in magnitude between ages 17 and 20. Likewise, some node pairs, such as GAD and ADHD, were found to have stagnant edges between ages 14 and 17 that sharply increased by age 20. These edge fluctuations were in turn verified by the NCTs, which revealed significant differences in both the structure and global connectivity of the networks across time. Structural differences were most pronounced for ages 14 and 20 ($M = 0.22$, $p < .001$), trailed by ages 14 and 17 ($M = 0.20$, $p = .002$), and 17 and 20 ($M = 0.15$, $p = .04$). In contrast, significant differences in global connectivity were largest between ages 14 and

17 ($S = 0.58$, $p < .001$), followed by ages 14 and 20 ($S = 0.30$, $p < .01$), and ages 17 and 20 ($S = 0.29$, $p < .001$). Significant differences in edge weights were most common for ADHD, CD/ASPD, MDD, GAD, and substance use related nodes (see Table S7 in the supplementals). In combination, these results suggested that the overall connectivity of the network was characterized by both decreases and increases, with the structure of the network similarly changing over time.

Discussion

The present study evaluated the explanatory power of dynamic mutualism theory in accounting for the developmental trajectories of the p -factor from ages 14–21. As research has remained limited in documenting the longitudinal trajectories of p from adolescence to young adulthood, we extend previous work by exploring the development and stability of p and the internalizing-externalizing factors during this important transitional period.

In efforts to provide a more comprehensive and theoretically compatible test of mutualism theory, we constructed three distinct statistical models to evaluate whether mutualistic processes were supported at the between- or within-person levels. Taken together, the results of the present study offer some support for the role of

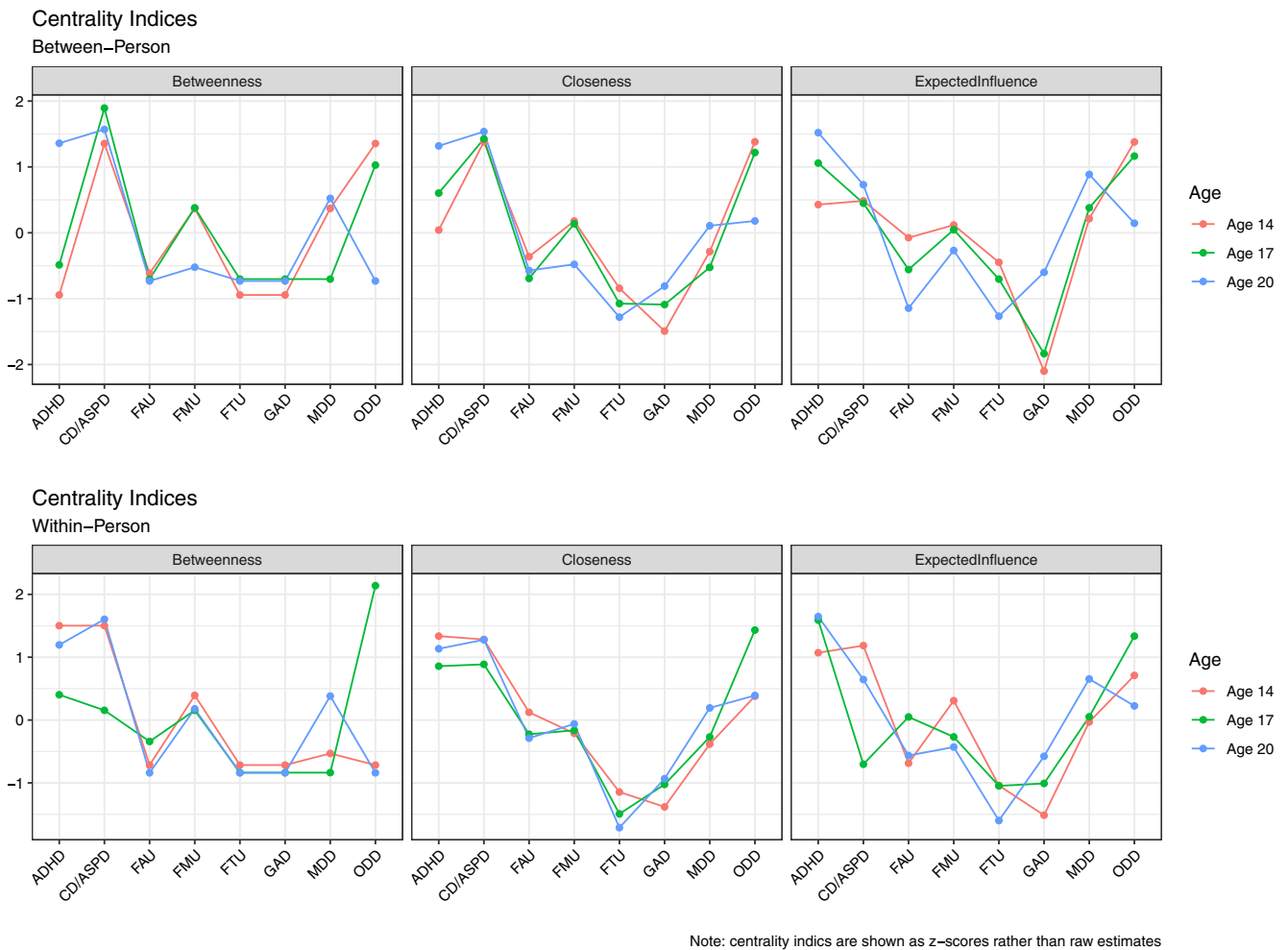


Figure 2. Centrality for between- and within-person networks by age. ADHD = attention-deficit hyperactivity disorder; CD/ASPD = conduct disorder/antisocial personality disorder traits; FAU = frequency of alcohol use; FMU = frequency of marijuana use; FTU = frequency of tobacco use; GAD = generalized anxiety disorder; MDD = major depressive disorder; ODD = oppositional defiant disorder.

mutualistic processes in the development of p ; however, our findings are intended to be preliminary in nature, and do not discount the potential for other processes or mechanisms to influence the development of psychopathology.

Regarding mutualistic processes at the between-person level, the bifactor models found support for a robust general p -factor that systematically increased in strength and variance explained with age. Wald tests indicated that these increases were unlikely due to chance, and that p and the internalizing-externalizing factors may be characterized by more nuanced dynamics from mid-adolescence to adulthood than previously reported (Castellanos-Ryan et al., 2016; Murray et al., 2016; Snyder et al., 2017). Furthermore, results indicated that the specific factors positively predicted p at several ages, with the inclusion of these cross-lagged paths significantly improving goodness of fit. Cross-lag effects tended to strengthen with age, particularly for internalizing predicting p and less so for externalizing and p . Consistent with mutualism theory, this suggested that symptom expression in a specific area of psychopathology (e.g., internalizing: depression) was associated with greater risk for developing broader symptoms from either domain in the future. Relatedly, our results found the between-person components of the internalizing and externalizing factors to be strongly correlated in the

RI-CLPMs. In line with the shared-risk hypothesis (Angold et al., 1999), this indicated that the co-development of internalizing and externalizing symptoms and related behaviors were partially attributable to stable, time-invariant risk factors that are shared across domains.

In terms of within-person associations, we found significant, albeit small, cross-lags between the internalizing and externalizing factors in the bidirectional RI-CLPM. That is to say that after accounting for the other between- and within-person effects, significant cross-lags still emerged, with the bidirectional model statistically outperforming other models. Results indicated that internalizing positively predicted externalizing starting at age 18, though these paths were only significant at ages 20 and 21. Clinically, these findings imply that targeting internalizing symptoms in late adolescence may be effective in preventing co-occurring externalizing problems from developing in adulthood.

In comparison, externalizing was a significant, though negative, predictor of internalizing at several ages throughout adolescence. This suggested that relatively higher levels of externalizing were associated with subsequent decreases in internalizing, highlighting a potential protective effect of externalizing in adolescence. Notably, these negative associations are inconsistent with longitudinal evidence that has reported externalizing to positively

predict within-person change in internalizing during early childhood (Oh et al., 2020), and at several points from childhood into adolescence (Flouri et al., 2019). Nonetheless, findings in this area have been mixed and other within-person studies have found externalizing to positively predict internalizing throughout childhood but negatively predict internalizing by early adolescence (Murray et al., 2020; Obsuth et al., 2020). Considering we examined these associations beginning at age 14, it is possible that our results reflect different developmental processes between adolescence and preceding stages of development, such that the positive associations predicted for internalizing and externalizing may be more pronounced in childhood rather than adolescence (e.g., Flouri et al., 2019; Murray et al., 2020; Oh et al., 2020).

Notably, when between- and within-person variance components were not directly separated, negative cross-lagged paths from externalizing to internalizing became less frequent, while positive cross-lags from internalizing to externalizing increased. These results highlight the importance of disentangling more stable between-person processes from within-person dynamics (Hamaker et al., 2015), and are congruent with studies that suggest within-person continuities between internalizing and externalizing are weaker once shared, time-invariant factors are controlled for at the between-person level (Wichstrøm et al., 2017).

Despite some negative associations in the RI-CLPMs, between- and within-person networks found several positive associations between internalizing and externalizing nodes over time. SWI estimates also increased with age and exceeded proposed cutoffs ($SWI \geq 1$), suggesting that the network structure became more densely connected throughout development. However, parallel to findings of Sterba et al. (2010), the estimated networks did not indicate a clear pattern of increasing or decreasing associations between internalizing and externalizing nodes. For example, while some internalizing and externalizing pairs increased in strength (e.g., ADHD and MDD), other node associations became weaker over time (e.g., ODD and MDD, FAU and MDD). Edges between ADHD and GAD were one of the few exceptions and remained more consistent, which may shed light on the positive cross-lags found for internalizing predicting externalizing in the RI-CLPM. Several studies have reported similar associations between internalizing symptoms and ADHD (Biederman et al., 2008; McElroy, Shevlin, et al., 2018; Murray et al., 2022; Speyer et al., 2021; Wichstrøm et al., 2017), though ADHD is usually suggested to increase the probability of anxiety and depression rather than the reverse. Yet, it is possible that after accounting for any stable commonalities between ADHD and GAD (i.e., deficits in executive functioning; Mogg et al., 2015), and ADHD and MDD (i.e., genetic overlap; Riglin et al., 2020), the effect of internalizing on ADHD symptoms is more identifiable at the within-person level during this developmental period (Murray et al., 2022).

In sum, the results discussed insofar provide some support for mutualism theory; however, it is important to note that some findings were also inconsistent with dynamic mutualism. First, the bifactor model indicated that p was a significant predictor of externalizing in late adolescence, and to a lesser degree, internalizing, which was incongruent with our original predictions. Given that p was largely defined by high levels of impulsivity/disinhibition (i.e., ADHD and ODD indicators) and the externalizing factor was mostly characterized by substance use and CD/ASPD traits, one interpretation of this finding is that it reflects links between impulsivity and substance use that are commonly found during mid-adolescence (Gullo & Dawe, 2008; Quinn & Harden, 2013), especially for girls (Kong et al., 2013). Conversely, this may suggest

that the relationship between p and externalizing during mid-adolescence is better characterized by differentiation-related processes rather than mutualistic processes, at least at the between-person level.

In addition to findings from the bifactor models, the bidirectional RI-CLPM suggested that a large proportion of significant cross-lags between internalizing and externalizing were negative. While these negative effects weakened over time, the fact that earlier periods of development were characterized by greater negative associations is largely inconsistent with a core prediction of dynamic mutualism (i.e., mostly positive associations). Considering several of our results were congruent with mutualism theory, it is possible that some of these mixed findings could reflect a misalignment between the measurements in the current study and the period in which these temporal dynamics truly unfold (Aristodemou et al., 2021). For example, causal interactions between internalizing and externalizing symptoms and/or disorders may be inadequately captured by the present study if these dynamics were stronger before age 14 or after age 21. Equally, it is conceivable that our yearly assessments may fail to capture some of the developing associations between internalizing and externalizing if these dynamics are better represented by more frequent measurements (e.g., weekly, monthly, bi-annually).

Alternatively, and akin to prior conclusions (Aristodemou et al., 2021; McElroy, Belsky, et al., 2018; Murray et al., 2020), these conflicting findings could alternatively be interpreted as evidence for multiple developmental processes to influence the overall trajectories of psychopathology. Seeing as the transition from adolescence to young adulthood is characterized by a multitude of biological, environmental, social, and psychological changes (Feldman et al., 1990; Masten, 2006), it seems reasonable that both mutualistic (e.g., narrow symptom presentations lead to broader expressions of psychopathology), and differentiation processes (e.g., broader forms of psychopathology lead to specific symptom manifestations) could describe the progression of psychopathology at different developmental stages. These underlying processes, in turn, may be influenced by other developmentally relevant distal or proximal factors that were unable to be incorporated in the current models (McLaughlin, 2016). For example, use of alcohol or other drugs among close friends is a robust predictor of adolescent substance use (Cousijn et al., 2018; Glaser et al., 2010), and if included in our analyses, may have led to slightly different conclusions.

Indeed, the potential for external factors to influence the measured trajectories of psychopathology was indirectly supported by the bidirectional RI-CLPM, which found large, positive within-time correlations between the internalizing and externalizing factors that increased over time. This suggested that girls who reported relatively higher levels of symptoms in one domain were found to report similar elevations in the other domain (and vice versa). These correlations could be interpreted as synchronous effects, or as evidence for unmeasured, time-variant factors that influence within-person change in both domains simultaneously (Willard et al., 2021). In other words, while the random-intercept factors control for the effect of stable, time-invariant influences – whether measured or unmeasured (Usami et al., 2019) – within-person relationships can still be confounded by external variables that change over time (Mund et al., 2021).

Assuming that time-varying factors exist and significantly influence the development of psychopathology, then the more pressing question becomes what these factors represent, and how such effects may emerge and change throughout development

(Kan *et al.*, 2019). Addressing this question from a dynamic mutualism perspective may argue that these effects are due to mutualistic processes that were not accounted for in our current models. For instance, numerous studies have documented links between poor executive functioning (EF) and psychopathology (Castellanos-Ryan *et al.*, 2016; M. M. Martel *et al.*, 2017; Moore *et al.*, 2020; Wade *et al.*, 2019), with a recent study suggesting that the development of poor EF and general psychopathology may be adequately described by a mutualism model, such that the effects of low EF compound throughout development and increase risk for developing multiple mental disorders in the future (Romer & Pizzagalli, 2021). Instead, and generally consistent with a common cause view of p , these effects could imply that after accounting for the more stable, communal features that p is often purported to reflect, other time-varying factors may still influence the expression of internalizing and externalizing symptoms to some degree. These time-varying effects could encompass a wide range of factors, extending from peer groups to parenting (Wichstrøm *et al.*, 2017), to developmental genetic changes, that, if present, may impact within-person parameters (Mund *et al.*, 2021; Pingault *et al.*, 2015).

These considerations once again underscore the notion that more than one developmental process likely shapes the trajectories of psychopathology, with different developmental periods characterized by varying mechanisms or processes (e.g., McElroy, Belsky, *et al.*, 2018; Wade *et al.*, 2019). For instance, causal mechanisms in childhood that are distinct or embedded within a specific latent vulnerability (e.g., internalizing) may confer greater risk for developing broader expressions of psychopathology in adolescence by activating a small range of symptoms that slowly expand via interactions with each other, and/or through interactions with other genetic or environmental factors (Lahey *et al.*, 2021). In adulthood, these symptom dynamics may be characterized by an entirely different process such as differentiation, in which symptoms become increasingly narrow with age (Lahey *et al.*, 2004; Lilienfeld *et al.*, 1994). Despite the simplistic nature of this scenario, it provides one example of how aspects of mutualism, differentiation, and elements of common cause positions can be combined to better understand or investigate possible accounts of development.

The attention to possible causal agents and the subsequent interplay between local-level symptoms also aligns with other recent frameworks for studying psychopathology, such as multi-causal (Kendler, 2019), or hybrid modeling approaches (e.g., Borsboom *et al.*, 2019; Bringmann & Eronen, 2018; Fried & Cramer, 2017; Koss & Gunnar, 2018). Although definitions of hybrid models vary to some extent, Fried and Cramer (2017) proposed a hybrid modeling framework that integrates traditional network and latent variable techniques to unveil how common causes and direct symptom interactions may cooperatively influence the onset and maintenance of a disorder, respectively. Otherwise stated, causal factors (represented as latent variables) are theorized to initiate the expression of certain symptoms, which, in turn, may facilitate the development of new symptoms via local interactions (Fried & Cramer, 2017).

For example, the death of a relative (i.e., a hypothesized causal variable) may lead to symptoms of depression in some individuals (e.g., loss of appetite, sleep disturbances), that subsequently enables the disordered state to be maintained over time by promoting the development of other symptoms (e.g., sleep deprivation leads to increased anxiety; Pires *et al.*, 2016). Future research may thus benefit from examining the extent to which different

developmental frameworks can account for more specific elements of psychopathology, as well as how these proposed mechanisms may change throughout the lifespan.

Notwithstanding the promise that hybrid models may hold for future research, it is worth noting that the success of these models critically relies on the ability of researchers to identify and measure potential causal mechanisms, which has proven to be challenging with respect to p (van Bork *et al.*, 2017). However, in cases where this may be feasible, residual network models (RNMs) provide one statistical solution for examining assumptions of hybrid models, as RNMs allow the network structure to be estimated after accounting for the influence that a latent variable has on its item covariances (Epskamp *et al.*, 2017). Though more research in this area is needed, a recent study examining childhood maltreatment and eating disorder symptoms also proposed a hybrid modeling approach that integrates network analysis and mediation models, highlighting other viable solutions (Monteleone *et al.*, 2019).

Strengths and Limitations

The racially diverse sample and longitudinal study design represent key strengths that enabled the development of p to be probed in participants often underrepresented in research and during a key transitional period not previously examined. Likewise, although prior dynamic mutualism studies used large community samples, they relied on parental or teacher reports and only one study was considered adequately diverse with respect to nationality, ethnicity, and socioeconomic status. Thus, the diversity of the current sample and use of self-report data provided the opportunity to compare how these developmental processes may have differed based on sample characteristics and type of informant report.

Another key strength of the present study was the integration of different statistical techniques, including a model that disaggregated between- and within-person processes (Hamaker *et al.*, 2015). While this differs from most p -factor studies that have used between-person techniques to study change, methods that attend to both intra- and interindividual trends are likely to yield more accurate descriptions of developing symptoms. Further, the combination of latent variable and network approaches may be especially fruitful for advancing research in this area, as the integration of techniques is more likely to describe the complexities of psychopathology than either approach alone (Eaton, 2015).

Despite these strengths, our study is subject to the following limitations. First, although our multi-method approach allowed for a more suitable test of mutualism theory compared to earlier studies, these analyses reflect a simplified version of the mutualism model of intelligence, and some assumptions of the theory (e.g., multiplier effects that are routed through the environment) were not directly tested. Likewise, given that mutualistic processes are not necessarily expected to increase during all points of development (Kievit, 2020), it is possible that our study failed to capture key dynamics between internalizing and externalizing that emerged during earlier stages of development. Second, our analyses were unable to incorporate other key forms of psychopathology (e.g., psychosis, personality disorders) that if included, may have led to different conclusions. In order to fully appreciate the development of co-occurring psychopathology, it will be critical for research to delineate changes in p when a more diverse array of symptoms are represented (Levin-Aspenson *et al.*, 2021; Shields *et al.*, 2020). Third, due to the

nature of our sample, the generalizability of these results may be limited, and replications in samples that are diverse with respect to sex and gender are encouraged. In the same vein, future replications in clinical or high-risk samples will be equally beneficial, as rates of psychopathology in the current sample are expected to be lower compared to clinical populations.

Fourth, our index of substance use was based on the average frequency of alcohol, marijuana, and tobacco use and did not consider the quantity of use. Therefore, results pertaining to substance use may be less generalizable and are unlikely to distinguish between more normative or experimental use of alcohol or other drugs as opposed to problematic substance use (Deas et al., 2000). Fifth, findings were based on self-reported measures of psychopathology, which may lead to inflated correlations due to common method variance (Richardson et al., 2009) or overlapping diagnostic criteria (e.g., trouble concentrating; Milberger et al., 1995). Sixth, our internalizing factor was comprised of only two indicators and would be considered under-identified if separated from the larger model. Relatedly, because this factor was constructed based on symptoms of generalized anxiety and depression, our representation of internalizing may not be comparable to studies that have included both distress and fear disorders (e.g., Gomez et al., 2019; Krueger & Markon, 2006). Sixth, network models were estimated cross-sectionally and cannot speak to whether bidirectional feedback loops or self-reinforcing edges were present in the network structure. Consequently, data with sufficient timepoints for estimating multilevel networks will be beneficial for discerning directionality and examining whether feedback loops are present between internalizing and externalizing indices (Epskamp, Waldorp, et al., 2018).

Conclusions

The p -factor is often conceptualized as an overarching predisposition to psychopathology, with on-going debate regarding a preferred statistical model for its study (Lahey et al., 2021). In exploring alternative theories of p and its development, the present study offers preliminary support for the use of a dynamic mutualism model in understanding the development of p and the internalizing-externalizing factors. In doing so, we hope to promote a more open dialogue surrounding the utility and substantive meaning of the p -factor, as well as encourage researchers to consider alternative frameworks and methodologies when investigating its development. Exploring different theoretical models may not only foster an increased understanding of the developmental mechanisms underlying p , but in turn may lead to novel insights into the prevention and treatment of psychopathology.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0954579422000463>

Acknowledgements. The authors greatly appreciate the families who took part in this study and the support from the PGS team, which includes interviewers and their supervisors, data managers, student workers, and volunteers.

Funding statement. This research was supported by grants from the National Institute of Mental Health (R01 MH56630, PI: Loeber; MH101088, PI: Stepp) and the National Institute on Drug Abuse (R01 DA012237, PI: Chung).

Conflicts of interest. The authors have declared no other financial disclosures or competing conflicts of interest.

References

- Aitken, M., Haltigan, J. D., Szatmari, P., Dubicka, B., Fonagy, P., Kelvin, R., ... Goodyer, I. M. (2020). Toward precision therapeutics: General and specific factors differentiate symptom change in depressed adolescents. *Journal of Child Psychology and Psychiatry*, 61(9), 998–1008. <https://doi.org/10.1111/jcpp.13194>
- Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(1), 57–87. <https://doi.org/10.1111/1469-7610.00424>
- Aristodemou, M. E., & Fried, E. I. (2020). Common factors and interpretation of the p factor of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 59(4), 465–466. <https://doi.org/10.1016/j.jaac.2019.07.953>
- Aristodemou, M. E., Kievit, R., Murray, A. L., Eisner, M., Ribeaud, D., & Fried, E. I. (2021). Common cause vs dynamic mutualism: An empirical comparison of two theories of psychopathology in two large longitudinal cohorts. *Mapping Intimacies*. <https://doi.org/10.31234/osf.io/a6ght>
- Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, 88(4), 1186–1206. <https://doi.org/10.1111/cdev.12660>
- Biederman, J., Ball, S. W., Monuteaux, M. C., Mick, E., Spencer, T. J., McCreary, M., ... Faraone, S. V. (2008). New insights into the comorbidity between ADHD and major depression in adolescent and young adult females. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(4), 426–434. <https://doi.org/10.1097/CHI.0b013e31816429d3>
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(4), 545–553. <https://doi.org/10.1097/00004583-199704000-00018>
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305. <https://doi.org/10.1037/0033-2909.110.2.305>
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. <https://doi.org/10.1177/2167702616657069>
- Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 42. <https://doi.org/10.1017/S0140525X17002266>
- Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One*, 6(11), e27407. <https://doi.org/10.1371/journal.pone.0027407>
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203. <https://doi.org/10.1037/0033-295X.110.2.203>
- Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The p factor and the n factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science*, 7(6), 1266–1284. <https://doi.org/10.1177/2167702619859332>
- Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review*, 125(4), 606. <https://doi.org/10.1037/rev0000108>
- Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, 49(5), 460–470. <https://doi.org/10.1080/00273171.2014.933697>
- Carragher, N., Teesson, M., Sunderland, M., Newton, N., Krueger, R., Conrod, P., ... Slade, T. (2016). The structure of adolescent psychopathology: A symptom-level analysis. *Psychological Medicine*, 46(5), 981–994. <https://doi.org/10.1017/S0033291715002470>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... Poulton, R. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. <https://doi.org/10.1177/2167702613497473>
- Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831–844. <https://doi.org/10.1176/appi.ajp.2018.17121383>

- Castellanos-Ryan, N., Brière, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., . . . Gallinat, J. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology, 125*(8), 1039. <https://doi.org/10.1037/abn0000193>
- Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology, 70*(1), 6. <https://doi.org/10.1037//0022-006x.70.1.6>
- Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., . . . Allison, E. (2019). Changes in general and specific psychopathology factors over a psychosocial intervention. *Journal of the American Academy of Child & Adolescent Psychiatry, 58*(8), 776–786. <https://doi.org/10.1016/j.jaac.2018.11.011>
- Conway, C. C., Mansolf, M., & Reise, S. P. (2019). Ecological validity of a quantitative classification system for mental illness in treatment-seeking adults. *Psychological Assessment, 31*(6), 730. <https://doi.org/10.1037/pas0000695>
- Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Möttus, R., Waldorp, L. J., & Cramer, A. O. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality, 54*, 13–29. <https://doi.org/10.1016/j.jrp.2014.07.003>
- Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences, 136*, 68–78. <https://doi.org/10.1016/j.paid.2017.06.011>
- Cousijn, J., Luijten, M., & Ewing, S. W. F. (2018). Adolescent resilience to addiction: A social plasticity hypothesis. *The Lancet Child & Adolescent Health, 2*(1), 69–78. [https://doi.org/10.1016/S2352-4642\(17\)30148-7](https://doi.org/10.1016/S2352-4642(17)30148-7)
- Dalsgaard, S., Thorsteinsson, E., Trabjerg, B. B., Schullehner, J., Planaripoll, O., Brikell, I., . . . Timmerman, A. (2020). Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence. *JAMA Psychiatry, 77*(2), 155–164. <https://doi.org/10.1001/jamapsychiatry.2019.3523>
- Deas, D., Riggs, P., Langenbucher, J., Goldman, M., & Brown, S. (2000). Adolescents are not adults: Developmental considerations in alcohol users. *Alcoholism: Clinical and Experimental Research, 24*(2), 232–237. <https://doi.org/10.1111/j.1530-0277.2000.tb04596.x>
- Deutz, M. H., Geeraerts, S. B., Belsky, J., Deković, M., van Baar, A. L., Prinzie, P., & Patalay, P. (2020). General psychopathology and dysregulation profile in a longitudinal community sample: Stability, antecedents and outcomes. *Child Psychiatry & Human Development, 51*(1), 114–126. <https://doi.org/10.1007/s10578-019-00916-2>
- Dueber, D. M. (2017). *Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models [Computer software]*. <https://doi.org/10.13023/edp.tool.01>
- Dunn, T. J., Baguley, T., & Brunson, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Eaton, N. R. (2015). *Latent variable and network models of comorbidity: Toward an empirically derived nosology*. Springer.
- Epskamp, S. (2015). *Bootnet: Bootstrap methods for various network estimation routines (R package version 0.2.) [Computer software]*. <https://CRAN.R-project.org/package=bootnet>
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212. <https://doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1–18. <https://doi.org/10.18637/jss.v048.i04>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*(4), 904–927. <https://doi.org/10.1007/s11336-017-9557-x>
- Epskamp, S., Waldorp, L. J., Möttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research, 53*(4), 453–480. <https://doi.org/10.1080/00273171.2018.1454823>
- Feldman, S. S., Elliott, G. R., & Elliott, G. R. (1990). *At the threshold: The developing adolescent*. Harvard University Press.
- Flouri, E., Papachristou, E., Midouhas, E., Ploubidis, G. B., Lewis, G., & Joshi, H. (2019). Developmental cascades of internalising symptoms, externalising problems and cognitive ability from early childhood to middle adolescence. *European Psychiatry, 57*, 61–69. <https://doi.org/10.1016/j.eurpsy.2018.12.005>
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286. <https://doi.org/10.1037/1040-3590.7.3.286>
- Forbes, M. K., Rapee, R. M., & Krueger, R. F. (2019). Opportunities for the prevention of mental disorders by reducing general psychopathology in early childhood. *Behaviour Research and Therapy, 119*, 103411. <https://doi.org/10.1016/j.brat.2019.103411>
- Foygel, R., & Drton, M. (2010). *Extended Bayesian information criteria for Gaussian graphical models* (pp. 604–612). <https://arxiv.org/abs/1011.6640>
- Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry, 31*(4), 271–288. <https://doi.org/10.1080/1047840X.2020.1853461>
- Fried, E. I., & Cramer, A. O. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science, 12*(6), 999–1020. <https://doi.org/10.1177/1745691617705892>
- Fried, E. I., Greene, A. L., & Eaton, N. R. (2021). The p factor is the sum of its parts, for now. *World Psychiatry, 20*(1), 69. <https://doi.org/10.1002/wps.20814>
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441. <https://doi.org/10.1093/biostatistics/kxm045>
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164. <https://doi.org/10.1002/spe.4380211102>
- Gadow, K., & Sprafkin, J. (1999). *Youth's inventory-4 manual*. Checkmate Plus.
- Gadow, K., Sprafkin, J., & Weiss, M. (2004). *Adult self-report inventory-4 manual*. Checkmate Plus.
- Glaser, B., Shelton, K. H., & van den Bree, M. B. (2010). The moderating role of close friends in the relationship between conduct problems and adolescent substance use. *Journal of Adolescent Health, 47*(1), 35–42. <https://doi.org/10.1016/j.jadohealth.2009.12.022>
- Gomez, R., Stavropoulos, V., Vance, A., & Griffiths, M. D. (2019). Re-evaluation of the latent structure of common childhood disorders: Is there a general psychopathology factor (p-factor)? *International Journal of Mental Health and Addiction, 17*(2), 258–278. <https://doi.org/10.1007/s11469-018-0017-3>
- Greene, A. L., & Eaton, N. R. (2017). The temporal stability of the bifactor model of comorbidity: An examination of moderated continuity pathways. *Comprehensive Psychiatry, 72*, 74–82. <https://doi.org/10.1016/j.comppsy.2016.09.010>
- Gullo, M. J., & Dawe, S. (2008). Impulsivity and adolescent substance use: Rashly dismissed as “all-bad”? *Neuroscience & Biobehavioral Reviews, 32*(8), 1507–1518. <https://doi.org/10.1016/j.neubiorev.2008.06.003>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102. <https://doi.org/10.1037/a0038889>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In *Structural equation modeling: Present and future* (pp. 195–216). Scientific Software International.
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But . . . *Communication Methods and Measures, 14*(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Hayward, C. (2003). *Gender differences at puberty*. Cambridge University Press.
- Hipwell, A. E., Loeber, R., Stouthamer-Loeber, M., Keenan, K., White, H. R., & Kroneman, L. (2002). Characteristics of girls with early onset disruptive and antisocial behaviour. *Criminal Behaviour and Mental Health, 12*(1), 99–118. <https://doi.org/10.1002/cbm.489>
- Hofman, A., Kievit, R., Stevenson, C., Molenaar, D., Visser, I., & van der Maas, H. (2018). The dynamics of the development of mathematics skills:

- A comparison of theories of developing intelligence [Unpublished manuscript]. <https://doi.org/10.31219/osf.io/xa2ft>
- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Humphries, M. D., & Gurney, K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One*, 3(4), e0002051. <https://doi.org/10.1371/journal.pone.0002051>
- Hygen, B. W., Skalická, V., Stenseng, F., Belsky, J., Steinsbekk, S., & Wichstrøm, L. (2020). The co-occurrence between symptoms of internet gaming disorder and psychiatric disorders in childhood and adolescence: Prospective relations or common causes? *Journal of Child Psychology and Psychiatry*, 61(8), 890–898. <https://doi.org/10.1111/jcpp.13289>
- Isvoranu, A.-M., & Epskamp, S. (2021). Continuous and ordered categorical data in network psychometrics: Which estimation method to choose? *Deriving Guidelines for Applied Researchers*. <https://doi.org/10.31234/osf.io/mbycn>
- Kan, K. J., van der Maas, H. L., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, 73, 52–62. <https://doi.org/10.1016/j.intell.2018.12.004>
- Keenan, K., Hipwell, A., Chung, T., Stepp, S., Stouthamer-Loeber, M., Loeber, R., & McTigue, K. (2010). The Pittsburgh Girls Study: Overview and initial findings. *Journal of Clinical Child & Adolescent Psychology*, 39(4), 506–521. <https://doi.org/10.1080/15374416.2010.486320>
- Kendler, K. S. (2019). From many to one to many—the search for causes of psychiatric illness. *JAMA Psychiatry*. <https://doi.org/10.1001/jama.psychiatry.2019.1200>
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602. <https://doi.org/10.1001/archpsyc.62.6.593>
- Kievit, R. A., Hofman, A. D., & Nation, K. (2019). Mutualistic coupling between vocabulary and reasoning in young children: A replication and extension of the study by Kievit et al. (2017). *Psychological Science*, 30(8), 1245–1252. <https://doi.org/10.1177/0956797619841265>
- Kievit, R. A. (2020). Sensitive periods in cognitive development: A mutualistic perspective. *Current Opinion in Behavioral Sciences*, 36, 144–149. <https://doi.org/10.1016/j.cobeha.2020.10.007>
- Kong, G., Smith, A. E., McMahon, T. J., Cavallo, D. A., Schepis, T. S., Desai, R. A., . . . Krishnan-Sarin, S. (2013). Pubertal status, sensation-seeking, impulsivity, and substance use in high-school-aged boys and girls. *Journal of Addiction Medicine*, 7(2), 116. <https://doi.org/10.1097/ADM.0b013e31828230ca>
- Koss, K. J., & Gunnar, M. R. (2018). Annual research review: Early adversity, the hypothalamic-pituitary-adrenocortical axis, and child psychopathology. *Journal of Child Psychology and Psychiatry*, 59(4), 327–346. <https://doi.org/10.1111/jcpp.12784>
- Kramer, M. D., Krueger, R. F., & Hicks, B. M. (2008). The role of internalizing and externalizing liability factors in accounting for gender differences in the prevalence of common psychopathological syndromes. *Psychological Medicine*, 38(1), 51–61. <https://doi.org/10.1017/S0033291707001572>
- Kramer, T. L., Phillips, S. D., Hargis, M. B., Miller, T. L., Burns, B. J., & Robbins, J. M. (2004). Disagreement between parent and adolescent reports of functional impairment. *Journal of Child Psychology and Psychiatry*, 45(2), 248–259. <https://doi.org/10.1111/j.1469-7610.2004.00217.x>
- Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*, 107(2), 216. <https://doi.org/10.1037//0021-843x.107.2.216>
- Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2, 111–133. <https://doi.org/10.1146/annurev.clinpsy.2.022305.095213>
- Laceulle, O. M., Vollebergh, W. A., & Ormel, J. (2015). The structure of psychopathology in adolescence: Replication of a general psychopathology factor in the TRAILS study. *Clinical Psychological Science*, 3(6), 850–860. <https://doi.org/10.1177/2167702614560750>
- Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, 121(4), 971. <https://doi.org/10.1037/a0028355>
- Lahey, B. B., Applegate, B., Waldman, I. D., Loft, J. D., Hankin, B. L., & Rick, J. (2004). The structure of child and adolescent psychopathology: Generating new hypotheses. *Journal of Abnormal Psychology*, 113(3), 358. <https://doi.org/10.1037/0021-843X.113.3.358>
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). Validity and utility of the general factor of psychopathology. *World Psychiatry*, 16(2), 142–144. <https://doi.org/10.1002/wps.20410>
- Lahey, B. B., Moore, T. M., Kaczurkin, 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., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry*, 56(4), 415–422. <https://doi.org/10.1111/jcpp.12300>
- Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4), 649–662. <https://doi.org/10.1080/10705511.2011.607721>
- Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.
- Levin-Aspenson, H. F., Watson, D., Clark, L. A., & Zimmerman, M. (2021). What is the general factor of psychopathology? Consistency of the p factor across samples. *Assessment*, 28(4), 1035–1049. <https://doi.org/10.1177/1073191120954921>
- Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology: Science and Practice*, 1(1), 71. <https://doi.org/10.1111/j.1468-2850.1994.tb00007.x>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85–110.
- Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., . . . Bressan, R. 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. <https://doi.org/10.1037/abn0000205>
- Martel, M. (2013). Sexual selection and sex differences in the prevalence of developmental psychopathology: Childhood externalizing and adolescent internalizing disorders. *Psychological Bulletin*, 139(6), 1221–1259. <https://doi.org/10.1037/a0032247>
- Masten, A. S. (2006). Developmental psychopathology: Pathways to the future. *International Journal of Behavioral Development*, 30(1), 47–54. <https://doi.org/10.1177/0165025406059974>
- McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum Associates.
- McElroy, E., Belsky, J., Carragher, N., Fearon, P., & Patalay, P. (2018). Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: Dynamic mutualism or p-differentiation? *Journal of Child Psychology and Psychiatry*, 59(6), 667–675. <https://doi.org/10.1111/jcpp.12849>
- McElroy, E., Shevlin, M., Murphy, J., & McBride, O. (2018). Co-occurring internalizing and externalizing psychopathology in childhood and adolescence: A network approach. *European Child & Adolescent Psychiatry*, 27(11), 1449–1457. <https://doi.org/10.1007/s00787-018-1128-x>
- McLaughlin, K. A. (2016). Future directions in childhood adversity and youth psychopathology. *Journal of Clinical Child & Adolescent Psychology*, 45(3), 361–382. <https://doi.org/10.1080/15374416.2015.1110823>
- McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*, 86, 95–104. <https://doi.org/10.1016/j.brat.2016.06.006>
- Milberger, S., Biederman, J., Faraone, S. V., Murphy, J., & Tsuang, M. T. (1995). Attention deficit hyperactivity disorder and comorbid disorder:

- Issues of overlapping symptoms. *The American Journal of Psychiatry*, 152(12), 1793–1799. <https://doi.org/10.1176/ajp.152.12.1793>
- Mogg, K., Salum, G., Bradley, B., Gadelha, A., Pan, P., Alvarenga, P., ... Manfro, G. (2015). Attention network functioning in children with anxiety disorders, attention-deficit/hyperactivity disorder and non-clinical anxiety. *Psychological Medicine*, 45(12), 2633. <https://doi.org/10.1017/S0033291715000586>
- Monteleone, A. M., Cascino, G., Pellegrino, F., Ruzzi, V., Patriciello, G., Marone, L., ... Maj, M. (2019). The association between childhood maltreatment and eating disorder psychopathology: A mixed-model investigation. *European Psychiatry*, 61, 111–118. <https://doi.org/10.1016/j.eurpsy.2019.08.002>
- Moore, T. M., Kaczurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., ... Kardan, O. (2020). Criterion validity and relationships between alternative hierarchical dimensional models of general and specific psychopathology. *Journal of Abnormal Psychology*, 129(7), 677. <https://doi.org/10.1037/abn0000601>
- Morin, A. J., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In *Handbook of sport psychology* (pp. 1044–1073). Wiley. <https://doi.org/10.1002/9781119568124.ch51>
- Mund, M., Johnson, M. D., & Nestler, S. (2021). Changes in size and interpretation of parameter estimates in within-person models in the presence of time-invariant and time-varying covariates. *Frontiers in Psychology*, 12, 666928. <https://doi.org/10.3389/fpsyg.2021.666928>
- Murray, A. L., Caye, A., McKenzie, K., Auyeung, B., Murray, G., Ribeaud, D., ... Eisner, M. (2022). Reciprocal developmental relations between ADHD and anxiety in adolescence: A within-person longitudinal analysis of commonly co-occurring symptoms. *Journal of Attention Disorders*, 26(1), 109–118. <https://doi.org/10.1177/1087054720908333>
- Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The development of the general factor of psychopathology ‘p factor’ through childhood and adolescence. *Journal of Abnormal Child Psychology*, 44(8), 1573–1586. <https://doi.org/10.1007/s10802-016-0132-1>
- Murray, A. L., Eisner, M., & Ribeaud, D. (2020). Within-person analysis of developmental cascades between externalising and internalising problems. *Journal of Child Psychology and Psychiatry*, 61(6), 681–688. <https://doi.org/10.1111/jcpp.13150>
- Newman, M. (2010). *Networks: An introduction*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- Newman, D. L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology*, 107(2), 305. <https://doi.org/10.1037//0021-843x.107.2.305>
- Obsuth, I., Murray, A. L., Di Folco, S., Ribeaud, D., & Eisner, M. (2020). Patterns of homotypic and heterotypic continuity between ADHD symptoms, externalising and internalising problems from age 7 to 15. *Journal of Abnormal Child Psychology*, 48(2), 223–236. <https://doi.org/10.1007/s10802-019-00592-9>
- Oh, Y., Greenberg, M. T., & Willoughby, M. T. (2020). Examining longitudinal associations between externalizing and internalizing behavior problems at within-and between-child levels. *Journal of Abnormal Child Psychology*, 48(4), 467–480. <https://doi.org/10.1007/s10802-019-00614-6>
- Olino, T. M., Bufferd, S. J., Dougherty, L. R., Dyson, M. W., Carlson, G. A., & Klein, D. N. (2018). The development of latent dimensions of psychopathology across early childhood: Stability of dimensions and moderators of change. *Journal of Abnormal Child Psychology*, 46(7), 1373–1383. <https://doi.org/10.1007/s10802-018-0398-6>
- Oltmanns, J. R., Smith, G. T., Oltmanns, T. F., & Widiger, T. A. (2018). General factors of psychopathology, personality, and personality disorder: Cross domain comparisons. *Clinical Psychological Science*, 6(4), 581–589. <https://doi.org/10.1177/2167702617750150>
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245–251. <https://doi.org/10.1016/j.socnet.2010.03.006>
- Pandina, R. J., Labouvie, E. W., & White, H. R. (1984). Potential contributions of the life span developmental approach to the study of adolescent alcohol and drug use: The Rutgers Health and Human Development Project, a working model. *Journal of Drug Issues*, 14(2), 253–268. <https://doi.org/10.1177/002204268401400206>
- Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry*, 207(1), 15–22. <https://doi.org/10.1192/bjp.bp.114.149591>
- Pettersson, E., Lahey, B. B., Larsson, H., & Lichtenstein, P. (2018). Criterion validity and utility of the general factor of psychopathology in childhood: Predictive associations with independently measured severe adverse mental health outcomes in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(6), 372–383. <https://doi.org/10.1016/j.jaac.2017.12.016>
- Pingault, J.-B., Rijdsdijk, F., Zheng, Y., Plomin, R., & Viding, E. (2015). Developmentally dynamic genome: Evidence of genetic influences on increases and decreases in conduct problems from early childhood to adolescence. *Scientific Reports*, 5(1), 1–9. <https://doi.org/10.1038/srep10053>
- Pires, G. N., Bezerra, A. G., Tufik, S., & Andersen, M. L. (2016). Effects of acute sleep deprivation on state anxiety levels: A systematic review and meta-analysis. *Sleep Medicine*, 24, 109–118. <https://doi.org/10.1016/j.sleep.2016.07.019>
- Quinn, P. D., & Harden, K. P. (2013). Differential changes in impulsivity and sensation seeking and the escalation of substance use from adolescence to early adulthood. *Development and Psychopathology*, 25(1), 223–239. <https://doi.org/10.1017/S0954579412000284>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, 12(4), 762–800. <https://doi.org/10.1177/1094428109332834>
- Riglin, L., Leppert, B., Dardani, C., Thapar, A. K., Rice, F., O'Donovan, M. C., ... Thapar, A. (2020). ADHD and depression: Investigating a causal explanation. *Psychological Medicine*, 1–8. <https://doi.org/10.1017/S003291720000665>
- Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, 125(6), 747. <https://doi.org/10.1037/abn0000181>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137. <https://doi.org/10.1037/met0000045>
- Romer, A. L., & Pizzagalli, D. A. (2021). Is executive dysfunction a risk marker or consequence of psychopathology? A test of executive function as a prospective predictor and outcome of general psychopathology in the adolescent brain cognitive development study[®]. *Developmental Cognitive Neuroscience*, 51, 100994. <https://doi.org/10.1016/j.dcn.2021.100994>
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5-12 (BETA). *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66(1), 20–40. https://doi.org/10.1207/s15327752jpa6601_2
- Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, 39(3), 472–480. <https://doi.org/10.1037/0022-3514.39.3.472>
- Sallis, H., Szekely, E., Neumann, A., Jolicoeur-Martineau, A., Van IJzendoorn, M., Hillegers, M., ... Tiemeier, H. (2019). General

- psychopathology, internalising and externalising in children and functional outcomes in late adolescence. *Journal of Child Psychology and Psychiatry*, 60(11), 1183–1190. <https://doi.org/10.1111/jcpp.13067>
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.
- Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivariate Behavioral Research*, 53(3), 419–429. <https://doi.org/10.1080/00273171.2018.1455142>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schulenberg, J. E., Sameroff, A. J., & Cicchetti, D. (2004). The transition to adulthood as a critical juncture in the course of psychopathology and mental health. *Development and Psychopathology*, 16(4), 799–806. <https://doi.org/10.1017/S0954579404040015>
- Schulenberg, J. E., & Zarrett, N. R. (2006). Mental health during emerging adulthood: Continuity and discontinuity in courses, causes, and functions. In J. J. Arnett, & J. L. Tanner (Eds.), *Emerging adults in America: Coming of age in the 21st century* (pp. 135–172). American Psychological Association. <https://doi.org/10.1037/11381-006>
- Shields, A. N., Giljen, M., España, R. A., & Tackett, J. L. (2020). The p factor and dimensional structural models of youth personality pathology and psychopathology. *Current Opinion in Psychology*, 37, 21–25. <https://doi.org/10.1016/j.copsyc.2020.06.005>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107. <https://doi.org/10.1007/s11336-008-9101-0>
- Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The general factor of psychopathology. *Annual Review of Clinical Psychology*, 16, 75–98. <https://doi.org/10.1146/annurev-clinpsy-071119-115848>
- Snyder, H. R., Young, J. F., & Hankin, B. L. (2017). Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clinical Psychological Science*, 5(1), 98–110. <https://doi.org/10.1177/2167702616651076>
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292. <https://doi.org/10.1037/11491-006>
- Speyer, L. G., Eisner, M., Ribeaud, D., Luciano, M., Auyeung, B., & Murray, A. L. (2021). Developmental relations between internalising problems and ADHD in childhood: A symptom level perspective. *Research on Child and Adolescent Psychopathology*, 49(12), 1567–1579. <https://doi.org/10.1007/s10802-021-00856-3>
- Sterba, S. K., Copeland, W., Egger, H. L., Jane Costello, E., Erkanli, A., & Angold, A. (2010). Longitudinal dimensionality of adolescent psychopathology: Testing the differentiation hypothesis. *Journal of Child Psychology and Psychiatry*, 51(8), 871–884. <https://doi.org/10.1111/j.1469-7610.2010.02234.x>
- Stoel, R. D., Garre, F. G., Dolan, C., & Van Den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, 11(4), 439. <https://doi.org/10.1037/1082-989X.11.4.439>
- Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In *Handbook of item response theory modeling* (pp. 201–224). Routledge.
- Tackett, J. L., Lahey, B. B., Van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology*, 122(4), 1142. <https://doi.org/10.1037/a0034151>
- Ullsperger, J. M., & Nikolas, M. A. (2017). A meta-analytic review of the association between pubertal timing and psychopathology in adolescence: Are there sex differences in risk? *Psychological Bulletin*, 143(9), 903. <https://doi.org/10.1037/bul0000106>
- Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24(5), 637. <https://doi.org/10.1037/met0000210>
- van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, 27(6), 759–773. <https://doi.org/10.1177/0959354317737185>
- van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2016). Package 'NetworkComparisonTest'. <https://cran.r-project.org/web/packages/NetworkComparisonTest/index.html>
- van Borkulo, C. D., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2017). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. <https://doi.org/10.1037/met0000476>
- van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- van der Maas, H. L., Kan, K. J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence*, 5(2), 16. <https://doi.org/10.3390/jintelligence5020016>
- Van Roy, B., Groholt, B., Heyerdahl, S., & Clench-Aas, J. (2010). Understanding discrepancies in parent-child reporting of emotional and behavioural problems: Effects of relational and socio-demographic factors. *BMC Psychiatry*, 10(1), 1–12. <https://doi.org/10.1186/1471-244X-10-56>
- Wade, M., Zeanah, C. H., Fox, N. A., & Nelson, C. A. (2019). Global deficits in executive functioning are transdiagnostic mediators between severe childhood neglect and psychopathology in adolescence. *Psychological Medicine*, 50, 1687–1694. <https://doi.org/10.1017/S0033291719001764>
- Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological Inquiry*, 31(4), 310–320. <https://doi.org/10.1080/1047840X.2020.1853476>
- 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>
- Wichstrøm, L., Belsky, J., & Steinsbekk, S. (2017). Homotypic and heterotypic continuity of symptoms of psychiatric disorders from age 4 to 10 years: A dynamic panel model. *Journal of Child Psychology and Psychiatry*, 58(11), 1239–1247. <https://doi.org/10.1111/jcpp.12754>
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>
- Willard, J. A., Agache, A., Kohl, K., Bihler, L.-M., & Leyendecker, B. (2021). Longitudinal interrelations between nonword repetition and vocabulary from age three to five: Evidence for within-child processes? *Developmental Psychology*, 57(9), 1423. <https://doi.org/10.1037/dev0001230>
- Williams, D. R., & Rast, P. (2020). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*, 73(2), 187–212. <https://doi.org/10.1111/bmsp.12173>
- Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, 54(5), 719–750. <https://doi.org/10.1080/00273171.2019.1575716>
- Zarrett, N., & Eccles, J. (2006). The passage to adulthood: Challenges of late adolescence. *New Directions for Youth Development*, 206(111), 13–28. <https://doi.org/10.1002/yd.179>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>