CAMBRIDGE
UNIVERSITY PRESS

Regular Article

Internalizing symptoms, well-being, and correlates in adolescence: A multiverse exploration via cross-lagged panel network models

Louise Black , Margarita Panayiotou   and Neil Humphrey

Manchester Institute of Education, University of Manchester, Manchester, UK

Abstract

Internalizing symptoms are the most prevalent mental health problem in adolescents, with sharp increases seen, particularly for girls, and evidence that young people today report more problems than previous generations. It is therefore critical to measure and monitor these states on a large scale and consider correlates. We used novel panel network methodology to explore relationships between internalizing symptoms, well-being, and inter/intrapersonal indicators. A multiverse design was used with 32 conditions to consider the stability of results across arbitrary researcher decisions in a large community sample over three years ($N = 15,843$, aged 11–12 at Time 1). Networks were consistently similar for girls and boys. Stable trait-like effects within anxiety, attentional, and social indicators were found. Within-person networks were densely connected and suggested mental health and inter/intrapersonal correlates related to one another in similar complex ways. The multiverse design suggested the particular operationalization of items can substantially influence conclusions. Nevertheless, indicators such as thinking clearly, unhappiness, dealing with stress, and worry showed more consistent centrality, suggesting these indicators may play particularly important roles in the development of mental health in adolescence.

Keywords: adolescence, mental health, network psychometrics, well-being

(Received 1 September 2020; revised 28 January 2021; accepted 9 March 2021; First Published online 15 June 2021)

Adolescence is recognized as a key developmental phase characterized by rapid physical, social, and psychological change (Dahl, Allen, Wilbrecht, & Suleiman, 2018; Patton et al., 2016; Sawyer, Azzopardi, Wickremarathne, & Patton, 2018). The majority of lifetime disorders also show first onset in the teenage years (Jones, 2013). Early adolescence is likely particularly important to understanding what sets in motion changes in mental health. For instance, key gender differences emerge, and contextual factors such as puberty and school transition are in process (Patalay & Fitzsimons, 2017, 2018). Evidence of the correlates of mental health in this age could therefore be key to improving identification, intervention, and prevention (Patalay & Fitzsimons, 2018). However, based on available evidence, methodological challenges make it difficult to determine which indicators are particularly important (see sections outlining analytical considerations below). This study therefore makes use of a new panel network model (Epskamp, 2020b) to consider indicator-level interactions (pairwise causal associations that are often bidirectional; Epskamp, Rhemtulla, & Borsboom, 2017) between internalizing symptoms, well-being, and inter/intrapersonal indicators.

Sharp increases in internalizing problems, particularly for girls, make up much of the mental health difficulties faced by adolescents (Rapee et al., 2019), and evidence suggests levels of these problems are increasing over time (Collishaw, 2015; NHS Digital, 2018). Furthermore, internalizing problems are also highly comorbid with other disorders (Carrellas, Biederman, & Uchida, 2017; Merikangas et al., 2010; NHS Digital, 2018; Wolff & Ollendick, 2006) and substantially correlated with other symptoms (Black, Panayiotou, & Humphrey, 2019; Patalay et al., 2015), making them an important focus for inquiry. Large numbers of adolescents also experience subthreshold internalizing symptoms. For instance, up to 12% of 11–14-year-olds experience subthreshold levels of depression (Bertha & Balázs, 2013). Here we consider the key theoretical and analytical considerations in the robust study of internalizing problems.

Theoretical considerations

The role of well-being

Further insight into internalizing problems and those at risk might be afforded by also measuring well-being (Bartels, Cacioppo, van Beijsterveldt, & Boomsma, 2013). This reflects the World Health Organization's longstanding definition that mental health should not consist only of the absence of symptoms (WHO, 1946). This broader conceptualization is also likely to be more useful in nonclinical samples, since positive mental health can capture greater variability (Alexander, Salum, Swanson, & Milham, 2020). Well-being is also closely related to internalizing symptoms, statistically and conceptually (Black, Panayiotou, & Humphrey, 2020b), showing substantial correlations for total

scores and latent constructs (.41–.68; Antaramian, Huebner, Hills, & Valois, 2010; Black et al., 2019; Suldo, Thalji, & Ferron, 2011). Correlations around this level suggest that constructs are substantially related while each still contributes distinct information.

Furthermore, given that self-report adolescent mental health problem data can be error-prone, it can be argued that well-being might be used to strengthen measurement. Specifically, substantial measurement error in adolescent mental health problems is suggested by low inter-rater associations and varying approaches to classification, and there is no clear criterion against which such measures can be validated (Wolpert & Rutter, 2018). Commonly used symptom measures are typically old and/or based on limited psychometric investigation (Bentley, Hartley, & Bucci, 2019; Black, Mansfield, & Panayiotou, 2020a; Dedrick, Greenbaum, Friedman, Wetherington, & Knoff, 1997; Goodman, 2001), whereas newer well-being measures that followed modern and rigorous item-development and validation standards (e.g., Ravens-Sieberer et al., 2005; Stewart-Brown et al., 2009), may complement symptom data and improve measurement accuracy. Routine adoption of such measures is also empirically justified since well-being seems to relate at a similar level to different domains of psychopathology as these relate to one another (e.g., Black et al., 2019). Since these psychopathology domains have been amalgamated into composites (e.g., Patalay & Fitzsimons, 2016), and there is conceptual and statistical similarity at the indicator level for internalizing symptoms and well-being (Black et al., 2020b), using well-being measures to capture additional information can be a useful approach.

Intra and interpersonal correlates of internalizing symptoms in adolescence

There is a substantial body of literature covering the developmental risk and promotive correlates of mental health in adolescents (for reviews see for example, Evans, Li, & Whipple, 2013; Fritz, de Graaff, Caisley, van Harmelen, & Wilkinson, 2018; Masten & Barnes, 2018). Moreover, there is theoretical consensus that systems models are appropriate (Bronfenbrenner, 2005; Evans et al., 2013; Masten & Barnes, 2018), namely considering factors from across personal (e.g., problem solving), family (e.g., secure attachment), and wider environments (e.g., school connectedness), and key correlates have consistently been identified across samples and methods (Masten & Barnes, 2018). It is important to capture these multiple systems since effects can cascade from one level to the other such that the interaction between mental health and environments is inherently complex (Masten & Cicchetti, 2010). Consistent with other literature considering the dynamic interplay of correlates and mental health, we focus on malleable (i.e., intra and interpersonal factors) rather than biological or socioeconomic variables (Fritz et al., 2019).

For internalizing problems in adolescence specifically, it is thought that social factors and emotional regulation are particularly key factors (Rapee et al., 2019), suggesting these should be particularly studied in the development of internalizing symptoms. The sudden physical, psychological, and social changes experienced in adolescence might affect expectations and views of young people, and these changes likely in turn impact internalizing symptoms (Rapee et al., 2019). Perceived home, peer, and school support are therefore likely important correlates. More generally, emotion regulation can be impacted by difficult home environments (e.g., maternal depression or parental conflict), and resulting difficulties managing emotions pose significant risk for internalizing problems (Thompson, 2019).

Gender differences

Inclusion of such correlates also facilitates consideration of a key issue for internalizing symptoms in adolescence, namely that these disproportionately affect girls (Merikangas et al., 2010; NHS Digital, 2018), and that this is increasingly the case (Bor, Dean, Najman, & Hayatbakhsh, 2014; Collishaw, 2015). A key theme in the theoretical literature is whether girls and boys experience quantitatively or qualitatively different risk factors (Hyde, Mezulis, & Abramson, 2008). Indicator-level analysis of internalizing symptoms, well-being, and relevant malleable correlates over time may therefore shed light on this question. For instance, it may be that previous construct-level analyses have made differences difficult to pin-point with variation occurring (qualitatively) at the indicator level. Alternatively, if a common network structure that varies in edge strength is found, quantitative differences may explain prevalence findings.

Analytical considerations

Within- and between-person effects

Longitudinal data consist of both variation within individuals (over time), and variation between individuals (Curran & Bauer, 2011). In panel data, people are nested in time, much as in multilevel data, for instance, children are nested in schools. This allows for the consideration of how variables influence one another within people on average over time, taking account of stable (or trait-like) individual differences. In the estimation of the cross-lagged panel network, estimated stable means, and deviations from these over time allow for a network of trait-like effects over time, a longitudinal network of malleable effects over time, and a contemporaneous network of (undirected) state-like effects that happen within the lag considered (in our case more quickly than once a year). For example, adolescents' general tendencies to report anxiety might be related to their general tendencies to report perceived social support. This trait-like effect therefore needs to be controlled for when considering the direction and strength of the temporal association between anxiety and social support.

This kind of disaggregation has led to new findings in construct-level panel models. For instance, while bidirectional relationships have been observed for internalizing and externalizing symptoms, only the latter predicted the former when disaggregated effects were considered (Flouri et al., 2019; Oh et al., 2020). Similar findings have been observed for adolescent depression and self-esteem (Masselink et al., 2018). There is also early evidence in younger children that correlates at different ecological levels can interact reciprocally at the within-person level (after accounting for between-person effects). Kaufman, Kretschmer, Huitsing, and Veenstra (2020) found evidence of such effects for internalizing symptoms, parenting, and bullying.

Thus, to understand how temporal effects between psychological variables occur for the average individual, analysis of within-person effects, accounting for between-person differences, is needed. For instance, we might consider whether change in internalizing problems is predicted by bullying. Without disaggregated analysis, and assuming other requirements for causal inference are met (Rohrer, 2018), we cannot be sure that those experiencing symptoms are not in fact also those commonly targeted by bullies (a between-person effect). Crucially, while it is well established that disaggregation of within and between-person effects is needed for accurate inferences to be made, it is still common-place to assume within-person processes from analyses representing a

blend of within and between variance (Hamaker, Kuiper, & Grasman, 2015).

Network analysis

While the studies cited above have modeled within and between-person effects separately, they have relied on total scores and latent factors which treat individual symptoms as indicators of a given mental state. While this approach can be statistically equivalent (Fried, 2020), we argue it is theoretically problematic, given the absence of external evidence for disorders, the likelihood that mental health states are contributed to by a constellation of biological and environmental factors, and the fact many disorders share indicators (Borsboom, Cramer, & Kalis, 2018; Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011). Network approaches might better capture the nuance and complexity likely to be present in adolescent mental health (Kalisch et al., 2019). These offer the opportunity to consider individual indicators and correlates as outcomes and predictors while accounting for all other indicators in the model (Epskamp, 2020b; Kalisch et al., 2019). For example, from the network perspective we can consider the unique association of bullying and worry, and in longitudinal networks we can also track direction. Mental states and their correlates can therefore be represented as dynamic with interactive indicators, such that, for instance, bullying leads to worry, which in turn leads to somatic symptoms, which in turn leads to unhappiness. The modeling of indicators, and not latent variables, within network models is arguably particularly appropriate for internalizing symptoms in adolescence since evidence suggests a lack of clear clustering into theoretical disorders (e.g., depression, anxiety) for this domain (McElroy & Patalay, 2019; McElroy, Fearon, Belsky, Fonagy, & Patalay, 2018).

Item-level differences in reporting have also been found in young adolescents for internalizing and well-being (Black et al., 2019). Similarly, analysis of adult samples suggest indicator-level analysis could be important to understanding gender differences. Fried, Nesse, Zivin, Guille, and Sen (2014) found men reported more suicidal ideation and psychomotor symptoms of depression in response to stress, while women reported more fatigue, appetite, and sleep problems. Within-person analysis at the indicator-level could therefore be key to improving understanding of the development of internalizing symptoms, including gender differences.

The current study

The current study aimed to explore indicator-level within and between-person associations for internalizing symptoms, well-being, and inter/intrapersonal correlates via novel panel network models (Epskamp, 2020b). A conceptual demonstration of the panel network model, is shown in Figure 1. This diagram is simplified to aid interpretation and therefore shows parameters for only two indicators, while in the current study 22 are included. The existence of large panel studies represents an opportunity to consider longitudinal indicator-level associations in rich datasets, in which within-person effects can be modeled (Curran & Bauer, 2011). We therefore conducted secondary analysis of a dataset designed to explore and test new ways to improve mental health and well-being of young people aged 10–16. The current study was based on existing data which we were familiar with the HeadStart (HS) evaluation (Deighton et al., 2019). Therefore, a multiverse approach in which multiple combinations of possible reasonable decisions are analyzed in parallel, was used to avoid

researcher degrees of freedom obscuring results (Simmons, Nelson, & Simonsohn, 2011; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016; Weston, Ritchie, Rohrer, & Przybylski, 2019).

Given gender differences for internalizing symptoms and personal and social resources (Rapee et al., 2019), we expected that associations in within and between models would be noninvariant across girls and boys. We hypothesized that irrespective of gender (a) social problems (e.g., being bullied) would show positive associations with internalizing symptoms and negative with well-being; (b) well-being and symptoms would be negatively associated; (c) intrapersonal factors (e.g., the ability to handle stress) would be negatively associated with symptoms and positively with well-being; (d) social support would be negatively associated with symptoms and positively associated with well-being. Given the lack of studies analyzing disaggregated models, we were unable to specify which effects would be observed at within or between-person levels. Finally, we explored which indicators were the most influential and the most predicted.

Method

Background and procedure

We undertook secondary analysis of data from three annual waves (2017–2019) collected from a longitudinal cohort study. The project from which data were drawn aims to explore and test new ways to improve mental health and well-being of young people aged 10–16 and prevent serious mental health issues from developing.

Ethical approval was granted by the UCL ethics committee (reference: 8097/003), and opt-out parental consent was given for adolescents to complete secure online surveys during the school day. Teachers read out an information sheet which emphasized pupils' confidentiality and right to withdraw. Socio-demographic data were drawn from the National Pupil Database.

Participants

Data were collected from 15,859 pupils in year seven (age 11–12) at Time 1, from 118 secondary schools in England (52.7% female). Given the focus of the project, the sample was not drawn to be representative: 35.4% had ever been eligible for free school meals at Time 1 compared to the national figure of 28.5% eligible in the previous six years (Department for Education, 2017a); 12.0% had special educational needs (national figure = 14.4%; Department for Education, 2017c); in terms of ethnicity, 74.2% were white (national figure = 75.2%), 9.3% were Asian (national figure = 10.7%), 5.7% were of Black origin (national figure = 5.6%), 4.0% were of mixed origin (national figure = 5.0%), .2% were Chinese (national figure = .45), while 1.6% were classified as any other ethnic group (national figure = 1.75), and 1.5% were unclassified (national figure = 1.5%; Department for Education, 2017b). Of this total sample, 16 were removed from the current study since they had missing data for all items included for analysis.

Item selection

The conceptual domains explored in the current study (based on the literature reviewed above and indicators available in the dataset at each time point) were: internalizing symptoms (including
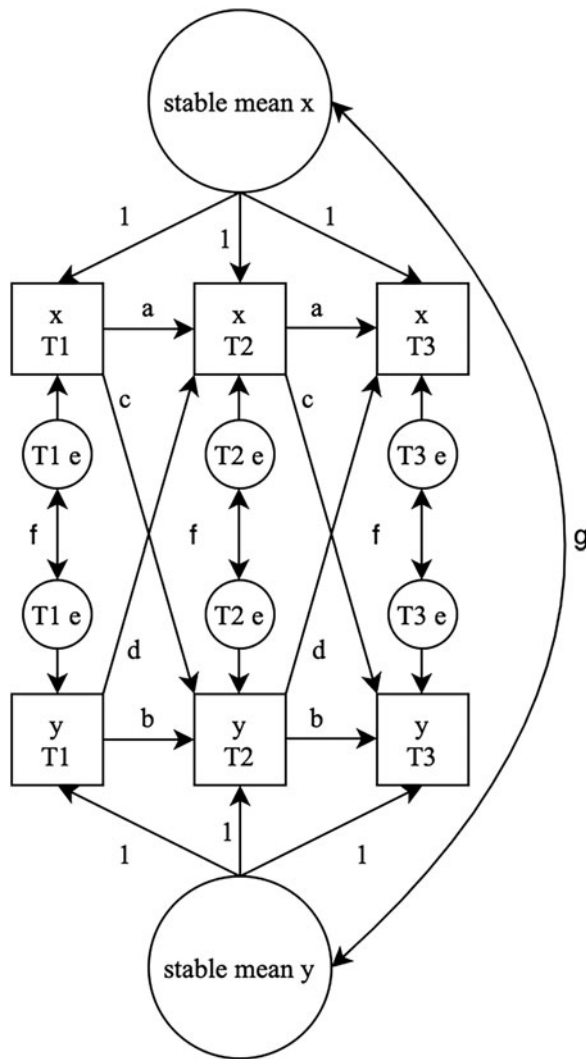
Figure 1. Conceptual diagram of a panel network model for two indicators, x and y, at three time points, T1–T3. Paths a–d represent average within-person directed partial correlations, including autocorrelations (temporal network). Paths marked f represent within-person partial correlations within lags (contemporaneous networks), with e representing the residual for each indicator after accounting for temporal effects. Path g represents between-person partial correlations for stable trait-like effects (between network).

attentional symptoms and social withdrawal; American Psychiatric Association, 2013; WHO, 2018), well-being, home, school and peer support, and intrapersonal factors such as managing stress. The choice of indicators was restricted, given that the software used for the panel network analysis currently cannot handle more than around 30 (Epskamp, 2020b) and it is not appropriate to indiscriminately include highly similar indicators in networks (Fried & Cramer, 2017; Rhemtulla, Cramer, van Bork, & Williams, 2018). Items were therefore selected from those available in the dataset according to the following criteria: (a) conceptual domain, (b) item simplicity, given issues highlighted in this area (Black et al., 2020a), (c) descriptive and factor model statistics.

The final list of items is shown in Table 1 alongside descriptive statistics (full item wording is available in the supplementary material, S1). Items were drawn from the Strengths and Difficulties Questionnaire (SDQ, Goodman, Meltzer, & Bailey,

1998), Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS, Stewart-Brown et al., 2009), Student Resilience Survey (SRS, Lereya et al., 2016), Trait Emotional Intelligence Questionnaire-Adolescent Short Form (TEIQUE-ASF, Petrides & Furnham, 2009), and four-item perceived stress scale (Demkowicz, Panayiotou, Ashworth, Humphrey, & Deighton, 2019).

Multiverse approach

In order to increase transparency, sensitivity analyses of many possible analytical decisions were conducted (Steegen et al., 2016; Weston et al., 2019). In line with multiple specification approaches (Simonsohn, Simmons, & Nelson, 2020; Steegen et al., 2016), variation in decisions was limited to those we considered likely to provide valid insight. In addition, given the novelty and computationally demanding nature of the analyses presented here, we also limited conditions based on feasibility. For instance, given that valid inferences can be drawn across a wide range of search algorithms at large sample sizes (Epskamp, 2020b), we used only two such robust, but relatively computationally light, procedures.[1]

Two aspects were identified as vulnerable to researcher degrees of freedom. First, the choice of items was in some cases arbitrary such that there were items from more than one scale that considered the same relevant experience. Second, the novelty of the method means that which estimation algorithm is most appropriate has not been clearly established. In such instances, multiverse approaches are recommended (Epskamp, 2019). This resulted in 16 possible datasets (based on varying two possible item operationalizations for four items: *distracted*, *mind*, *optimism*, *problem*, see Table 1) × two search algorithms, meaning that models for 32 conditions were estimated. For more details on the choice of items see the supplementary material (S1). Only full information maximum likelihood (FIML) estimation was selected since data cannot yet be treated as ordinal in the panel network model and no robust adjustments are available. Two equally robust pruning methods were considered: alpha at .01, and this plus stepwise modification based on the Bayesian information criterion (BIC).

We also kept the number of indicators in each model constant since we wanted to avoid overfitting by including multiple indicators of the same experience (e.g., two peer support items), and since networks are not directly comparable with varying numbers of nodes (Costantini et al., 2019). Missing data were retained for all conditions given FIML estimation, and since analysis was at the item level, and data were ordinal, outliers were not considered.

The resulting design allowed us to assess the stability of the most influential and predicted indicators and gender invariance across these decisions. Fit was assumed to be good across conditions given the data-driven approach, and was not used to compare conditions. Since our analysis was exploratory, testing multiple contingent effects, we approached the results of our sensitivity analyses descriptively in line with Steegen et al. (2016). We therefore present how fit, strength, and gender invariance varied across analyses.

Analysis

Code for all analyses, and simulated data for the purpose of running code, is available in the supplementary material (S2–S4). In

---

[1]We found it was not possible to run the modelsearch function with item-level panel analysis in *psychonetrics*.

Table 1. Node names, item wording and descriptive statistics over time

Node name/abbreviated wording	Measure	Mean (SD)	Skew
Bully/others bully me	SDQ	.31–.42 (.58–.66)	1.3–1.74
Close/feeling close to others	SWEMWBS	2.33–2.41 (1.10–1.15)	.51–.60
Distracted A/easily distracted	SDQ	1.01– 1.07 (.74–.76)	−.01–.11
Distracted B/my attention is good	SDQ	.76–.86 (.64)	.13–.25
Feelings/hard to control my feelings	TEIQUE-ASF	3.17–3.29 (1.9–2.0)	.47–.58
Help/when in need I find someone	SRS	2.38–2.75 (1.31–1.34)	.23–.58
Home/at home there is an adult who listens	SRS	1.76–1.87 (1.02–1.09)	1.14–1.31
Mind A/able to make up my own mind	SWEMWBS	2.10–2.25 (1.05)	.63–.79
Mind B/change my mind often	TEIQUE-ASF	3.17–.3.29 (1.90–2.0)	.47–.58
Nervous/nervous in new situations…easily lose confidence	SDQ	.99–1.09 (.76)	−.15–.02
Optimism A/optimistic about the future	SWEMWBS	2.64–2.69 (1.06–1.13)	.26–.33
Optimism B/things were going your way	PSS-4	1.93–1.98 (1.05–1.06)	.12–.14
Peer/there are students at school who make you feel better	SRS	1.93–2.02 (1.15–1.16)	.97–1.14
Problem A/dealing with problems well	SWEMWBS	2.64–2.74 (1.11–1.16)	.27–.32
Problem B/ability to handle personal problems	PSS-4	1.68–1.76 (1.13–1.19)	.23–.30
Relaxed/feeling relaxed	SWEMWBS	2.72–2.90 (1.11–1.12)	.03–.18
Restless/restless…cannot stay still	SDQ	1.10–1.12 (.73–.74)	−.15–.19
Scared/many fears…scared	SDQ	.61–.65 (.71–.72)	.65–.72
School/at school an adult really cares about me	SRS	2.46–2.74 (1.24–1.28)	.21–.45
Somatic/headaches, stomach-aches or sickness	SDQ	.74–.78 (.73–.76)	.40–.45
Stress/able to deal with stress	TEIQUE-ASF	3.72–3.85 (1.97–2.10)	.06–.14
Think/thinking clearly	SWEMWBS	2.47–2.70 (1.08)	.28–.41
Unhappy/unhappy, downhearted or tearful	SDQ	.53–.62 (.68–.71)	.70–.90
Useful/feeling useful	SWEMWBS	2.74–2.86 (1.04–1.07)	.11–.26
Withdrawn/usually on my own	SDQ	.43–.45 (.65–.67)	1.18–1.23
Worry/worry a lot	SDQ	.94–1.04 (.77–.79)	−.07–.11

Note. SDQ = strengths and difficulties questionnaire; SWEMWBS = short Warwick-Edinburgh mental well-being scale; TEIQUE-ASF = trait emotional intelligence questionnaire- adolescent short form; PSS-4 = 4-item perceived stress scale; SRS = student resilience survey.

order to count relationships counter to our hypotheses across conditions, all indicators were coded to have a positive manifold (e.g., well-being indicators were reversed with respect to symptoms). The first stage of the main analysis (for each condition) was to estimate a panel network model for each whole sample in the *psychonetrics* package in R (0.7.1; Epskamp, 2020a). Once the model was estimated, nonsignificant parameters were recursively pruned at α = .01 and then parameters were added one at a time based on modification indices to minimize the BIC, via the step-up function. This data-driven approach is consistent with network methods (Epskamp, Borsboom, & Fried, 2018a; Fried & Cramer, 2017). Given this, model fit was expected to be good, with comparative fit index (CFI) > .95 and root mean square error of approximation (RMSEA) < .06 (Hu & Bentler, 1999).

Once each full sample network was estimated (via basic pruning and stepwise modification), three matrices from the model were extracted for invariance testing and further consideration: (a) the temporal matrix which encodes directed partial correlations for the average within-person effects over time; (b) the contemporaneous matrix which encodes partial correlations for the average within-person effects within lags (after accounting for the temporal effects); (c) the between-persons matrix which encodes partial correlations for stable trait-like differences across all time points. Average networks across all 32 models, excluding edges that occurred less than 50% of the time following Lin, Fried, and Eaton (2020), were plotted in *qgraph* (1.6.5; Epskamp et al., 2012) with red lines indicating negative parameter values (edges) and blue positive. In the temporal network, arrows between nodes indicate directed partial correlations while curved arrows represent autoregressions.

Finally, strength centrality was considered for networks in each model. Strength represents the sum of absolute edge weights for any given node (Costantini et al., 2015). For temporal networks, this includes both in-strength and out-strength, with the former indicating the relative predictability and the latter the relative influence of the target node. For undirected networks, a single strength index represents the overall extent to which a given node is directly influenced by or influences others.

Network matrices were also inspected to determine the number and size of edges and whether these were in expected directions.

Gender invariance

Following standard practices for invariance testing two models were tested: an unconstrained model or H1, and a constrained model or H0, where H0 is nested in H1. Temporal, between, and contemporaneous matrices were used to determine which parameters should be considered in an unconstrained model (i.e., those retained in the whole sample were estimated for each group). In this model these parameters of interest were freely estimated in girls and boys simultaneously to provide a point of comparison for subsequent constraints. In the constrained model, all three matrices were then set to equality in girls and boys, and the resulting model was compared to the unconstrained model. Given the sample size of the current study, models were compared based on the Akaike information criterion (AIC) and BIC which penalize for model complexity (van de Schoot, Lugtig, & Hox, 2012), rather than chi-square difference testing which can be sensitive to large samples (Crede & Harms, 2019). Lower values for AIC and BIC indicate better model fit. Since the constrained model was more parsimonious, we interpreted higher AIC and BIC values for the constrained model as indicative of noninvariance.

Results

Gender was missing for .3% of the sample. Missing data for survey indicators were low for the first wave but higher for subsequent time points (Time 1 = 2.6%–6.6%, Time 2 = 16.4%–20.9%, Time 3 = 25.9%–29.6%). Descriptive statistics are summarized in Table 1. The average fit of models is presented in Table 2 and a full summary of fit statistics for each model can be found in the supplementary material (S5). In general, differences between equivalent datasets using different estimation algorithms were small indicating good stability across these. Though data-driven approaches were used to estimate models, and parameter estimates varied, the stable good fit across conditions nevertheless indicated that stationarity constraints imposed in the model (paths a and b in Figure 1) were reasonable in all cases (Epskamp, 2020b). In terms of invariance, the same mixed result was found across all conditions: AIC favored the unconstrained model while BIC favored the constrained model. This suggests differences in network structure between girls and boys were likely small. Post-hoc consideration of RMSEA and CFI also revealed differences typically considered to be small (Meade, Johnson, & Braddy, 2008),[2] (−.007 to −.006 for CFI, with $M = -.006$, $SD < .001$; range within $<-.001$ for RMSEA, $M < .001$, $SD < .001$).

Edges for contemporaneous and between networks are interpreted as partial correlation coefficients, and those for temporal as directed partial directed correlations (standardized beta coefficients). For each network within each condition the number of parameters, means, standard deviations, and number of negative edges (unexpected results relative to our hypotheses, given the recoding of indicators to have a positive manifold) can be seen in the supplementary material (S6). Between networks had the fewest edges (3–23), though these were relatively large (ranging in absolute value from $r = .004$–$r > .99$ for similar indicators such as *distracted* and *restless* in some models with the mean of mean edge sizes within networks across conditions $M = .40$). No unexpected negative edges were found for between networks in

any condition. Contemporaneous networks were more densely connected with 116–130 edges ($r = .06$–.27 in absolute value with the mean of mean edge sizes within networks across conditions $M = .07$), and with consistent unexpected negatives across all conditions (7–12; for example, a small negative edge featured in every contemporaneous network for the being bullied and [not] think clearly indicators). Temporal networks were also dense (166–196 edges; $\beta = .05$–.27 in absolute value, mean of means $M = .06$) with 1–3 unexpected edges found for each condition. These were consistently found for *worry → think*, *school → withdrawn*, and *unhappy → peer* (this was nonsignificant in eight conditions). Most estimated parameters were significant across conditions (0–12 were nonsignificant for any given condition, $p < .01$). In terms of edge parameters, only temporal networks occasionally included nonsignificant edges: 19 different edges in temporal networks were nonsignificant in different conditions, with most of these edges not occurring frequently across conditions or only rarely being nonsignificant (full information for all parameters in all conditions is provided in the supplementary material (S7–S10).

Spearman correlations between weight matrices for networks of the same type (e.g., temporal or contemporaneous) were high: Between $M \rho = .82$ $SD \rho = .14$; contemporaneous $M \rho = .88$ $SD \rho = .06$; temporal $M \rho = .91$ $SD \rho = .03$, suggesting networks were similar across conditions (full correlation matrices can be seen in the supplementary material, S11).

To summarize these networks across all conditions, for each of between, contemporaneous, and temporal networks, the mean of edges was calculated after excluding those that appeared in less than 50% of conditions. This resulted in 182 edges (37.60% of all possible edges) being retained in the average temporal network, all of which appeared across all conditions. Similar stability was seen for the between and contemporaneous networks, with all edges estimated across conditions appearing in 50% or more conditions (between: six edges, 2.60%; contemporaneous: 134 edges, 58.01%). The mean edge size for the average between network was $r = .32$ (range = .52), $r = .06$, (range = .32) for the average contemporaneous network, and $r = .06$ (range = .29) for the average temporal network. Autoregressive effects were present for all nodes in the average temporal network and ranged from .06 to .24 ($M = .14$, $SD = .04$). The average networks are summarized in Figure 2 with the thickness of edges scaled across the three panels (i.e., it is equivalent across each plot), and the supplementary material (S12). As mentioned above, a handful of nonsignificant parameters were found in temporal networks, six of which appear in the averaged network (all were nonsignificant only once across the 32 conditions, except unhappy → peer as described above).

Strength centrality was calculated for temporal and contemporaneous networks only, given the sparsity found for the between networks. Which nodes were most central, tended to depend on the condition. In and out strength for the temporal networks are shown in Figures 3 and 4, while strength for the contemporaneous network is shown in Figure 5. *Stress* was consistently high for in-strength but other nodes varied substantially. *Stress* was again fairly consistently one of the most central for out-strength as was *worry*, though again substantial variation in out-strength was seen for most nodes. *Worry* and *think* were consistently the most central for strength in the contemporaneous network. Nodes that were represented by varying items, depending on the condition, often showed particular discrepancies for strength (e.g., *mind* in Figure 3). However, nodes with the same item across conditions also showed substantial variation (e.g., *worry*

2We are not aware of simulation work providing recommendations for the size of alternative fit index differences for network invariance and therefore provide this example for confirmatory factor analysis (which recommends CFI difference of <.002 to consider invariance) since it includes larger samples closest to that used here.

Table 2. Average fit across datasets by model type

Model type	χ^2 M (SD)	Min/max	df M (SD)	Min/max	CFI M (SD)	Min/max	RMSEA M (SD)	Min/max	AIC M (SD)	Min/max	BIC M (SD)	Min/max
Unpruned	7364.94 (314.00)	6902.29/ 8028.15	1265.00 (0.00)	1265/ 1265	0.979 (0.001)	0.976/ 0.980	0.017 (0.000)	0.017/ 0.018	2180850.22 (26288.05)	2144674.95/ 2218121.15	2188612.75 (26288.05)	2152437.48/ 2225883.68
Pruned												
(p = .01)	13282.74 (529.08)	12288.01/ 14383.54	1906.88(9.02)	1892/ 1919	0.960 (0.002)	0.956/ 0.964	0.019 (0.000)	0.019/ 0.020	2185484.26 (26280.05)	2148923.91/ 2223074.19	2188323.30 (26280.96)	2151854.03/ 2225866.25
Stepup	12576.36 (1046.90)	10332.04/ 13593.12	1901.44 (14.94)	1865/ 1917	0.963 (0.003)	0.958/ 0.969	0.019 (0.001)	0.017/ 0.020	2184788.76 (26008.62)	2148429.69/ 2222664.55	2187669.51 (26033.41)	2151367.49/ 2225471.95
Pruned unconstrained	15812.44 (472.83)	14947.65/ 16875.93	3813.75 (18.05)	3784/ 3838	0.957 (0.002)	0.953/ 0.960	0.020 (0.000)	0.019/ 0.021	2175247.69 (26200.58)	2139017.27/ 2212536.20	2180923.85 (26202.95)	2144875.54/ 2218118.42
Stepup unconstrained	15249.43(948.84)	13049.01/ 16226.61	3802.88 (29.88)	3730/ 3834	0.959 (0.003)	0.954/ 0.965	0.020 (0.001)	0.018/ 0.020	2174706.42 (25933.39)	2138772.35/ 2212236.48	2180465.97 (25984.64)	2144645.95/ 2217849.37
Pruned constrained	17808.36 (477.39)	16983.10/ 18881.08	4117.88 (9.02)	4103/ 4130	0.951 (0.002)	0.946/ 0.954	0.021 (0.000)	0.020/ 0.021	2176635.36 (26235.31)	2140363.03/ 2213952.87	2179979.52 (26236.43)	2143798.24/ 2217250.06
Stepup constrained	17264.63 (918.01)	15192.08/ 18246.28	4112.44 (14.94)	4076/ 4128	0.953 (0.003)	0.948/ 0.959	0.020 (0.001)	0.019/ 0.021	2176102.51 (25966.31)	2140120.64/ 2213664.02	2179488.36 (25991.75)	2143563.53/ 2216976.55

Note. df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion; M = mean; SD = standard deviation.
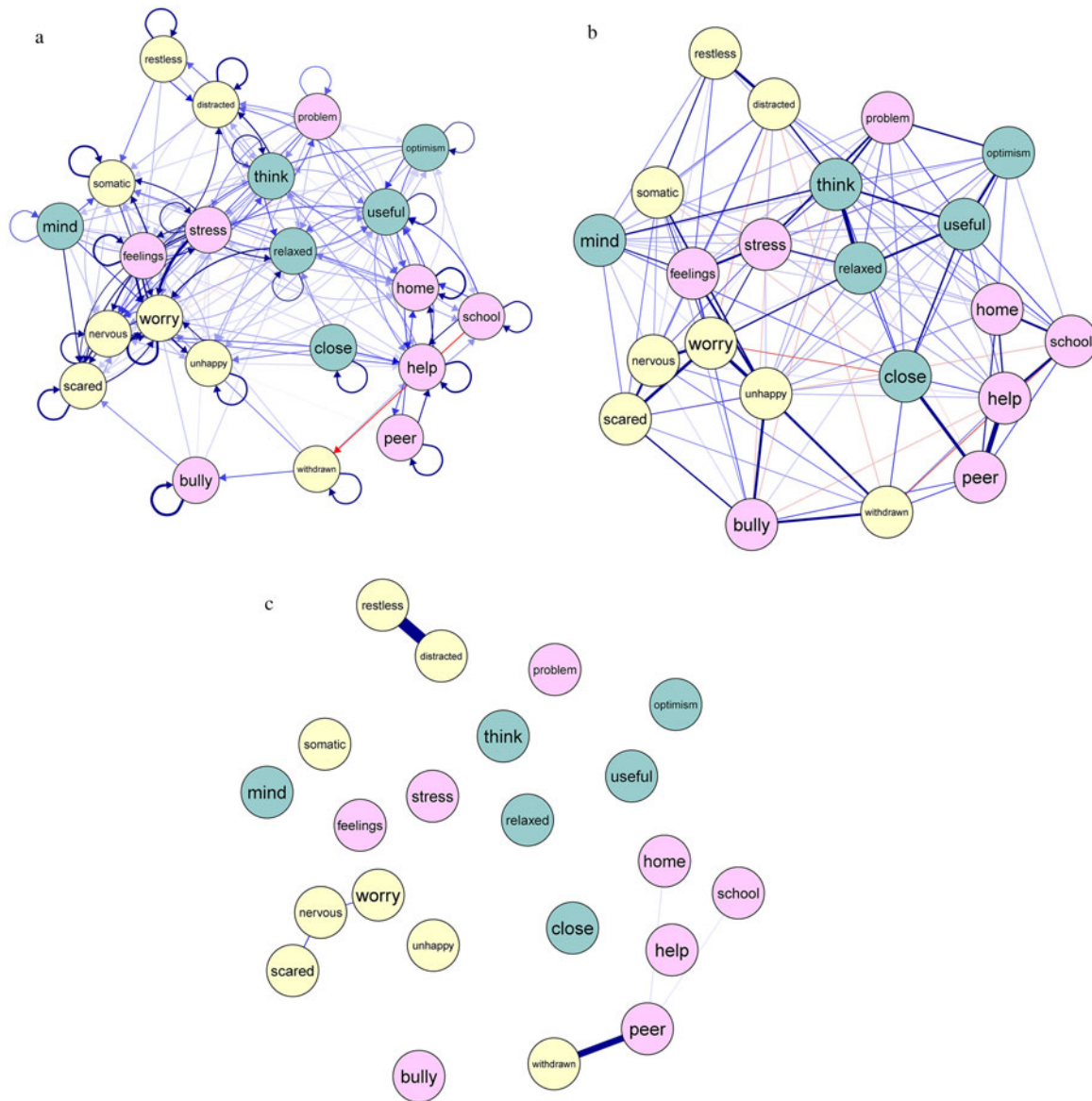
Figure 2. Average networks across conditions. Panel A: average temporal network. Panel B: average contemporaneous network. Panel C: average between network.

in Figure 3), given the conditional nature of edges which account for all others in the model.

Each line represents how the in-strength of each node varies depending on the condition. Only nodes where the maximum in-strength is always >.40 are shown in color and labeled for ease of reading.

Each line represents how the out-strength of each node varies depending on the condition. Only nodes where the maximum out-strength is always >.50 are shown in color and labeled for ease of reading.

Each line represents how the strength of each node varies depending on the condition. Only nodes where the maximum strength is always >.90 are shown in color and labeled for ease of reading.

Discussion

We explored stable trait-like and within-person associations over time for internalizing symptoms, well-being and inter and

intrapersonal correlates at the indicator level. A multiverse approach was adopted, varying estimation algorithms and operationalizations of certain indicators, given that secondary data analysis was conducted, and new methods were used (Epskamp, 2019; Weston et al., 2019). Though network analyses have boomed in recent years (Robinaugh, Hoekstra, Toner, & Borsboom, 2020), this was the first study to adopt a crossed multiverse design, to our knowledge, and an early example of Epskamp's (2020b) panel methodology. While previous work has considered longitudinal relationships between internalizing symptoms and inter/intrapersonal correlates (e.g., Goodman, Samek, Wilson, Iacono, & McGue, 2019; Saint-Georges & Vaillancourt, 2020), work at the indicator level was lacking. This revealed relationships between indicators of different domains, suggesting latent-variable approaches may miss complexity. Similarly, while some work has considered both symptoms and well-being over time (e.g., Patalay & Fitzsimons, 2018), covariance between these domains was only considered by controlling for each at the first time point.
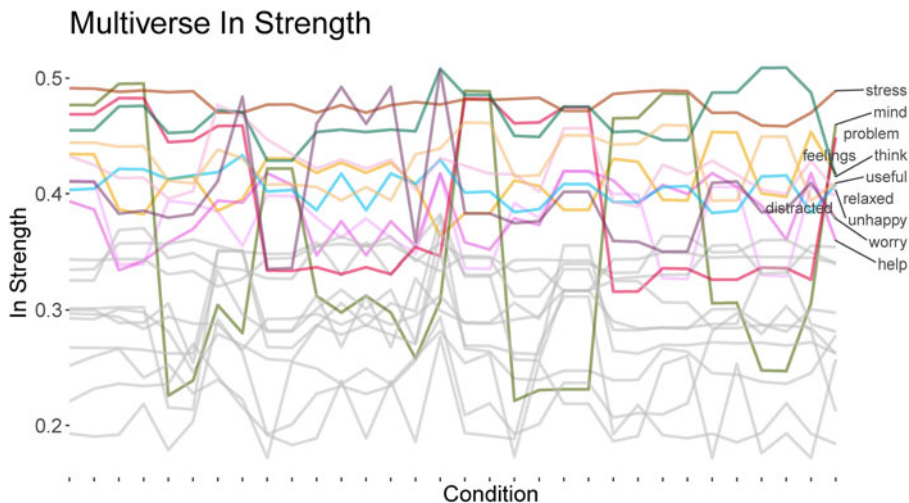
Multiverse In Strength

Figure 3. In-strength for each temporal network. Each line represents how the in-strength of each node varies depending on the condition. Only nodes where the maximum in-strength is always >.40 are shown in color and labeled for ease of reading.

We found a sparse between-person network with few strong associations, while the contemporaneous (average within-lag within-person associations) and temporal (directed within-person associations) networks, were densely connected. All weights matrices were highly correlated, and networks showed good stability across conditions. We did not find clear evidence that networks differed between girls and boys, and results were consistent across conditions. Findings suggest if differences existed for the indicators used here, they were likely trivial. Finally, the choice of item operationalization had a substantial impact on strength centrality (considered for the within-person networks), though certain nodes were consistently central.

Between-person Findings

The between network revealed partial correlations in expected directions, some of which were very large. These were between attentional, anxiety, and social indicators. These could reflect consistent cognitive vulnerabilities, environments, personality traits (e.g., agreeableness and neuroticism) or stable biological factors (Fraley & Roberts, 2005). There were notably no between-person relationships among indicators of different domains (e.g., internalizing and well-being or internalizing and social correlates) despite the fact that such domains have shown meaningful relationships elsewhere (e.g., Patalay & Fitzsimons, 2018). The contrast in our findings with prior work could result for several reasons, including our disaggregation of within and between-person effects, control of informant-type, and separate modeling of temporal and contemporaneous effects.

Though there was a relatively strong effect between peer support and withdrawal, which could be considered different domains (internalizing and interpersonal), we interpret this in line with the other effects in the between network: Those indicators involved were very similar and tended to be rated in similar ways over time, that is, a trait-like tendency over time to rate high or low peer support was strongly related to a trait-like tendency to rate low or high social withdrawal. The fact that indicators of different domains were conditionally independent in the between network suggests that covariance between these domains may be more state like. We were able to identify this by controlling for trait-like reporting effects in the between network. The relative sparsity of the between network also indicates the majority of covariances were not stable and trait-like, consistent with the rapidly changing

developmental context of early adolescence described in the introduction.

Within-person Findings

Dense within-person, temporal and contemporaneous, networks were found. These findings fit with systems approaches in which aspects from different levels (e.g., home and intrapersonal factors) interact with one another (Bronfenbrenner, 2005; Evans et al., 2013; Masten & Barnes, 2018). Furthermore, there was little evidence of particular associations for certain inter or intrapersonal factors being associated with only symptoms or well-being as has been suggested elsewhere (Patalay & Fitzsimons, 2016). Rather, symptoms, well-being and inter/intrapersonal factors seemed to influence one another in similar ways.

While both within-person networks were relatively dense, larger relationships were typically seen in the contemporaneous network. The current study sought to understand relationships between specific indicators (e.g., thinking clearly and being bullied) rather than latent constructs (e.g., well-being or peer problems). While levels of specific indicators such as these likely have meaningful relationships over time, the dense contemporaneous network suggests that interactions between the indicators modeled here often happened more quickly than annually (Epskamp et al., 2018a,b). Since both contemporaneous and temporal networks were relatively dense, many edges were common across both of these networks. Our results therefore suggest that indicators influenced one another both within and across lags. This further points to rapid changes in mental health and correlate variables, consistent with the rapid social, physical, and psychological development seen in early adolescence (Dahl et al., 2018; Patton et al., 2016; Sawyer et al., 2018). To better understand how these processes unfold, future work should vary the length between study waves, and there is a particular need for work focusing on shorter intervals.

Some of the larger effects in the temporal network were autoregressions, with each node showing such an effect. While the indicators studied here are known to be stable or show increasing trajectories (Meeus, 2016), meaning autoregressive effects would be expected, this finding is noteworthy. First, our analysis was at the indicator level, suggesting stability or reinforcement of these states can be specific to this level, rather than the domain (e.g., internalizing symptoms). Second, while latent-variable
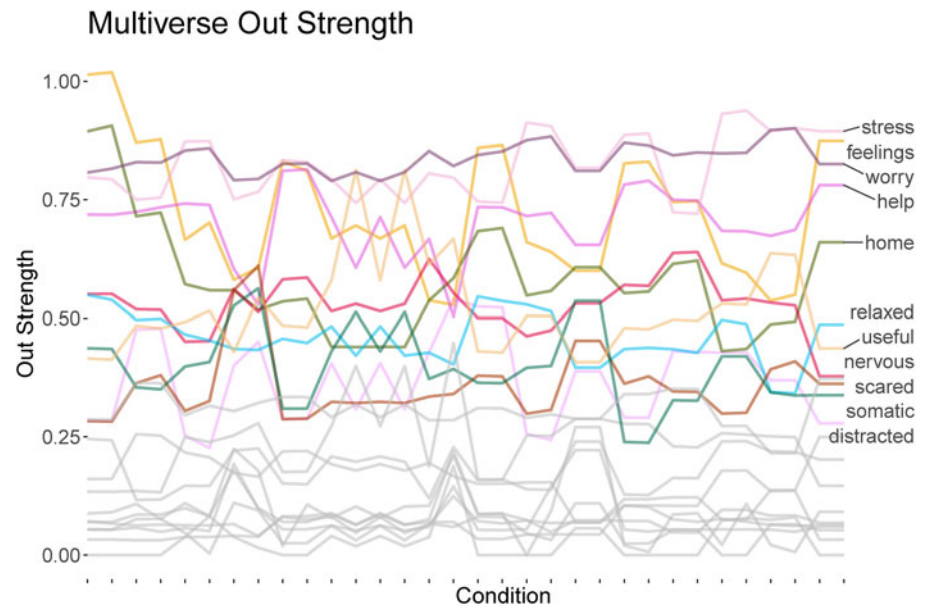
Multiverse Out Strength

Figure 4. Out-strength for each temporal network. Each line represents how the out-strength of each node varies depending on the condition. Only nodes where the maximum out-strength is always >.50 are shown in color and labeled for ease of reading.

approaches account for construct-level covariance, parameters in our model controlled for those to *all* other indicators, and thus also included unique variance beyond that explained by a potential latent variable, which could be substantively important. Third, autoregressive parameters in our analysis accounted for stable between-person differences over time, thus representing more accurate within-person reinforcement of individual experiences over time. These within-person autoregressions have been interpreted by some as warning signals for transition into more disordered states (e.g., van de Leemput et al., 2014), but since we examined a large cohort via survey methods, we did not consider whether individuals were more or less disordered over time.[3] Nevertheless, the age range studied here is thought to be critical in the emergence of mental health problems (Jones, 2013) and rates are known to increase in this age range (Merikangas et al., 2010; NHS Digital, 2018). It may be therefore that cementing of symptoms, well-being indicators and inter/intrapersonal factors all contribute to this change.

Edges were mostly in expected directions, relative to our hypotheses. However, unexpected negative parameters were observed consistently in the temporal and contemporaneous networks. A certain level of such effects in partial correlation networks could be consistent with the nominal alpha level, or due to conditioning on common effects (Epskamp, Waldorp, Mõttus, & Borsboom, 2018b). We are therefore cautious in providing substantive interpretation of these results. However, one such effect was particularly stable across conditions and temporal and contemporaneous networks, that between *withdrawn* and (lack of) school support.[4] While we anticipated that internalizing symptoms would be positively associated with perceived lack of social support, it may be that adolescents who reported feeling socially withdrawn were focusing on the peer level when responding to the *withdrawn* item. In fact, the full item reads "I am usually on my own. I generally play alone or keep to myself". It is

possible that adolescents who felt withdrawn from their peers tended to garner more support from, or were dependent on, school staff as can be the case for loneliness (Galanaki & Vassilopoulou, 2007).

Centrality

Strength centrality appeared more stable for the contemporaneous network than the temporal. *Think* and *unhappy* had the highest strength, depending on the condition, followed by *worry*, while the rank order of strength varied for the remaining nodes. This suggests that when considering relationships that happened more quickly than over a year, feeling unhappy and thinking clearly were particularly connected to other indicators, sharing the most variance with others (Costantini et al., 2015). Internalizing symptom and well-being indicators appeared to be among the most important in the contemporaneous network, suggesting both outcomes are intricately connected to each other and correlates. This further supports the use of well-being measures to better understand internalizing states, since well-being indicators clearly shared meaningful variance with other indicators without being redundant with respect to internalizing indicators.

Being able to deal with stress was one of the most consistently strongly predicted and influential nodes, suggesting that for effects that happened over the course of a year, the indicators in the model often related to this outcome via relatively strong directed partial correlations. Conversely, finding it hard to control feelings, an item designed to measure the same underlying trait as the *stress* indicator, was sometimes the most central for out-strength, while at other times several other indicators were stronger, and substantially lower values were seen. *Worry*, which was fairly consistently one of the most central nodes across conditions for out-strength, varied substantially for in-strength. Other particularly wide variations for in-strength were seen for the *mind* and *problem* indicators, both of which had varying operationalizations across conditions.

Given the finding that the contemporaneous network remained dense, we do not interpret only the temporal centrality results as indicative of risk factors or outcomes. Rather, results

---

3This would have relied on total scores which can be problematic (McNeish & Wolf, 2020) and inconsistent with our modeling approach.

4The node is considered as lack of school support due to the recoding prior to analysis to obtain a positive manifold for the easy detection of results counter to hypotheses across conditions.
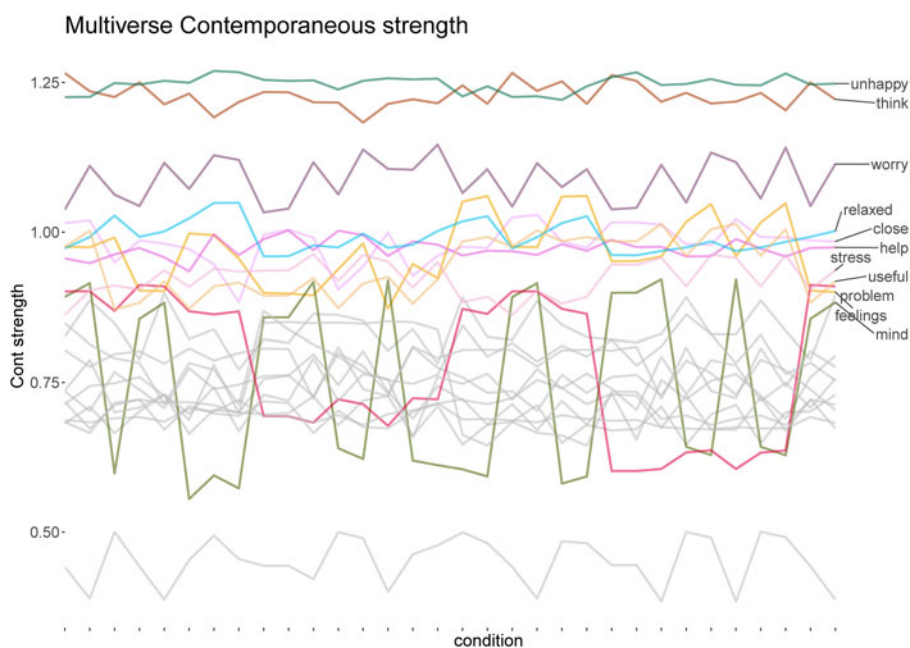
Figure 5. Contemporaneous strength. Each line represents how the strength of each node varies depending on the condition. Only nodes where the maximum strength is always >.90 are shown in color and labeled for ease of reading.

suggest that worry, managing stress, thinking clearly and unhappiness may be key indicators for the development of adolescents' mental health. While more work is needed, this suggests that worry and unhappiness may be particularly important symptoms in early adolescence when considering how rapid developmental change is navigated. In turn, the *think* and *stress* indicators' centrality suggest that such cognitive indicators may play an important role in the reinforcement of social and psychological processes in this age group.

Our findings also highlight the importance of which items are chosen, and the issues of measurement error in adolescent mental health data. The stability of networks across samples using the same items has been given attention in recent years (e.g., Borsboom et al., 2017; Forbes, Wright, Markon, & Krueger, 2017), as has the stability across different measures in certain fields (Fried et al., 2018). However, this was the first study, to our knowledge, to consider the sensitivity of network parameters to item operationalizations in the *same* sample. We found that while some nodes showed relative stability others varied in strength centrality for both indicators that were constant and those that varied across conditions.

Gender invariance

Gender invariance results were stable across conditions but, we were unable to determine clear support for invariance based on AIC and BIC as recommended by van de Schoot et al. (2012). Consistent with known possible behavior of these criteria, AIC favored the model with more parameters (unconstrained), while BIC did the opposite, favoring the constrained model (Vrieze, 2012). Since we had no clear rationale to favor one over the other, we consider these results in light of other literature and indices (post-hoc). Kan, van der Maas, and Levine (2019) found the same pattern of AIC and BIC for their unconstrained and constrained networks. They concluded that the same structure was applicable to both groups, though at least one edge varied in magnitude. Our post-hoc consideration of CFI and RMSEA also suggested trivial differences, thus supporting the approximate invariance of networks between boys and girls. Substantively, a single pattern of edges fitted both girls and boys, though small differences may exist in the strength of different relationships between particular nodes. Results should be replicated considering other measures and samples, but this suggests tentative evidence that girls and boys may experience quantitative rather than qualitative differences in risk and protective factors for internalizing symptoms, when considering inter/intrapersonal correlates.

Implications

Taken together, the strong dissociated relationships at the between level and densely connected nodes at the within level suggest the apparent discriminant validity of scales may particularly capture between-person differences rather than profiles within individuals. This is consistent with the fact that measures are typically developed using between-person (i.e., cross-sectional) data, such that the covariance structure from which the model is estimated describes variation between people (Molenaar, 2004). Though many analyses assume a blend of within and between effects is modeled without explicitly attending to this, it is often the case that within and between associations are not aligned (Curran & Bauer, 2011). Future work should therefore consider further the within and between properties of measures such as those used here, since they are typically used to probe within-person effects.

Findings further suggest integrated indicator-level approaches to adolescents' mental states and perceived resources should be considered, rather than testing to diagnose specific disorders. Our analysis therefore represents an example of how clinical and research approaches can better align, as has been pointed out for network approaches more generally (Borsboom, 2017). While formulations are often preferred over strict diagnostic criteria by clinicians (Johnstone, 2018), research has tended to rely on simplistic total scores or latent variables to define groups and categories. These are powerful approaches, with many advantages such as the estimation of measurement error. Nevertheless, as indicator-level approaches gain increasing attention (Robinaugh et al., 2020), analyses such as ours can offer more detailed insights. While much more work is needed, the current study demonstrates that brief surveys deployed in large samples

can be modeled in more nuanced ways. There is therefore potential to move beyond disorder-level (i.e., total-score) approaches. In addition, it may not be enough to disaggregate within and between effects at the construct level, since within-person effects likely happen across domains in a complex way (Borsboom et al., 2018). More transdiagnostic and indicator-level work is therefore needed to better understand within and between-person effects.

In addition to the substantive implications, our multiverse design revealed methodological issues. Where observed-data level networks are considered, as is typically the case (Robinaugh et al., 2020), rather than at the latent level (Epskamp, 2020b; Epskamp et al., 2017), researchers should be aware that item-level error may affect conclusions. Since many authors rely on single measures of each construct in their datasets, they will be unable to verify whether, for instance, centrality is robust to variations in items. Our out-strength results particularly demonstrate that had we chosen any one of the 32 conditions as the focus of our analysis, our conclusions could have varied substantially. While there are calls for increased use of latent networks, our approach also reveals that even in large rich datasets, there may not be enough indicators of each construct to conduct such analysis. For instance, our dataset had only one bullying item. We therefore echo the recent call for methods to be designed explicitly with network methods in mind (Robinaugh et al., 2020).

Strengths and limitations

The current study drew on a large sample, disaggregated within-person variance from stable trait-like effects, and incorporated a comprehensive multiverse design. Despite this a number of limitations must be acknowledged. First, the panel methodology adopted did not allow us to control for stable covariates, such as socioeconomic status. While the sample was purposively drawn to target those at risk, and therefore generally consisted of more deprived adolescents, there was variation in this. The sample was therefore also not representative and results should only be generalized to similar community samples with above average levels of deprivation.

We were also unable to account for the nonnormal ordinal nature of our data since this is not yet possible in *psychonetrics*. Nevertheless, this is consistent with much of the network literature to date, which often treats similar Likert-type data to that used here as continuous (Robinaugh et al., 2020). In addition, the use of a polychoric matrix to account for the ordinal nature of items in skewed data, such as that used here, can lead to bias (Fried, van Borkulo, & Epskamp, 2020). It can also lead to convergence issues in samples with substantial missingness, as was the case here, suggesting FIML was more appropriate. We also had little to draw on to interpret our invariance analyses, and more work is needed to understand the properties of fit indices when comparing networks.

Quality issues have been highlighted for some SDQ items, from which internalizing and bullying indicators were drawn (Black et al., 2020a), though self-report mental health measures are typically of low quality (Bentley et al., 2019). Finally, decisions about which items were interchangeable were subjectively considered based on content, though decisions in multiverse analyses are not expected to be uniform across researchers (Simonsohn et al., 2020).

Conclusion

The current multiverse panel network model allowed consideration of complex interactions between indicators of mental health and inter/intrapersonal factors consistent with theory and clinical approaches (Bronfenbrenner, 2005; Johnstone, 2018). Stable trait-like effects within anxiety, attentional and social indicators were found that were insensitive to analytical decisions. No clear differences were observed between boys and girls. Within-person networks were densely connected and relationships between indicators often unfolded within waves, suggesting more work should consider shorter lags. Mental health and inter/intrapersonal indicators appeared to relate to one another in similar complex ways. Our multiverse design revealed that the particular operationalization of items can have substantial effects on conclusions. Nevertheless, indicators such as thinking clearly, unhappiness, dealing with stress and worry showed more consistent centrality, suggesting these indicators may play particularly important roles in the development of mental health in adolescence.

References

Alexander, L. M., Salum, G. A., Swanson, J. M., & Milham, M. P. (2020). Measuring strengths and weaknesses in dimensional psychiatry. *Journal of Child Psychology and Psychiatry*, 61, 40–50. doi:10.1111/jcpp.13104

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th edn). Arlington, VA: American Psychiatric Association.

Antaramian, S. P., Huebner, S. E., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry*, 80, 462–472. doi:10.1111/j.1939-0025.2010.01049.x

Bartels, M., Cacioppo, J. T., van Beijsterveldt, T. C. E. M., & Boomsma, D. I. (2013). Exploring the association between well-being and psychopathology in adolescents. *Behavior Genetics*, 43, 177–190. doi:10.1007/s10519-013-9589-7

Bentley, N., Hartley, S., & Bucci, S. (2019). Systematic review of self-report measures of general mental health and wellbeing in adolescent mental

health. *Clinical Child and Family Psychology Review*, 22, 225–252. doi:10.1007/s10567-018-00273-x

Bertha, E. A., & Balázs, J. (2013). Subthreshold depression in adolescence: A systematic review. *European Child & Adolescent Psychiatry*, 22, 589–603. doi:10.1007/s00787-013-0411-0

Black, L., Mansfield, R., & Panayiotou, M. (2020a). Age appropriateness of the self-report strengths and difficulties questionnaire. *Assessment*, 0, 1073191120903382. doi:10.1177/1073191120903382

Black, L., Panayiotou, M., & Humphrey, N. (2019). The dimensionality and latent structure of mental health difficulties and wellbeing in early adolescence. *PLoS One*, 14, e0213018. doi:10.1371/journal.pone.0213018

Black, L., Panayiotou, M., & Humphrey, N. (2020b). The special relationship of internalizing symptoms and wellbeing: A cross-validation study considering indicator-level associations beyond the dual-factor model of mental health. doi:10.31234/osf.io/stajk

Bor, W., Dean, A. J., Najman, J., & Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Australian & New Zealand Journal of Psychiatry*, 48, 606–616. doi:10.1177/0004867414533834

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16, 5–13. doi:10.1002/wps.20375

Borsboom, D., Cramer, A., & Kalis, A. (2018). Brain disorders? Not really… Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 1–54. doi:10.1017/S0140525X17002266

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One*, 6, e27407. doi:10.1371/journal.pone.0027407

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, 126, 989–999. doi:10.1037/abn0000306

Bronfenbrenner, U. (2005). *Making human beings human: Bioecological perspectives on human development*. Thousand Oaks, USA: Sage.

Carrellas, N. W., Biederman, J., & Uchida, M. (2017). How prevalent and morbid are subthreshold manifestations of major depression in adolescents? A literature review. *Journal of Affective Disorders*, 210, 166–173. doi:10.1016/j.jad.2016.12.037

Collishaw, S. (2015). Annual research review: Secular trends in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*, 56, 370–393. doi:10.1111/jcpp.12372

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (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. doi: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. doi:10.1016/j.paid.2017.06.011

Crede, M., & Harms, P. (2019). Questionable research practices when using confirmatory factor analysis. *Journal of Managerial Psychology*, 34, 18–30. doi:10.1108/JMP-06-2018-0272

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62, 583–619. doi:10.1146/annurev.psych.093008.100356

Dahl, R. E., Allen, N. B., Wilbrecht, L., & Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature*, 554, 441. doi:10.1038/nature25770

Dedrick, R. F., Greenbaum, P. E., Friedman, R. M., Wetherington, C. M., & Knoff, H. M. (1997). Testing the structure of the child behavior checklist/4-18 using confirmatory factor analysis. *Educational and Psychological Measurement*, 57, 306–313. doi:10.1177/0013164497057002009

Deighton, J., Lereya, S., Casey, P., Patalay, P., Humphrey, N., & Wolpert, M. (2019). Prevalence of mental health problems in schools: Poverty and other risk factors among 28 000 adolescents in England. *British Journal of Psychiatry*, 215(3), 565–567. doi:10.1192/bjp.2019.19

Demkowicz, O., Panayiotou, M., Ashworth, E., Humphrey, N., & Deighton, J. (2019). The factor structure of the 4-item perceived stress scale in English

adolescents. *European Journal of Psychological Assessment*, 36, 913–917. doi:10.1027/1015-5759/a000562

Department for Education. (2017a). Pupil premium: allocations and conditions of grant 2016 to 2017. Retrieved from https://www.gov.uk/government/publications/pupil-premium-conditions-of-grant-2016-to-2017

Department for Education. (2017b). Schools, pupils and their characteristics: January 2017. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf

Department for Education. (2017c). Special educational needs in England: January 2017. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633031/SFR37_2017_Main_Text.pdf

Epskamp, S. (2019). Reproducibility and replicability in a fast-paced methodological world. *Advances in Methods and Practices in Psychological Science*, 2, 145–155. doi:10.1177/2515245919847421

Epskamp, S. (2020a). *Package 'psychonetrics'*. Retrieved from https://cran.r-project.org/web/packages/psychonetrics/psychonetrics.pdf

Epskamp, S. (2020b). Psychometric network models from time-series and panel data. *Psychometrika*, 85, 206–231. doi: 10.1007/s11336-020-09697-3

Epskamp, S., Borsboom, D., & Fried, E. I. (2018a). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50, 195–212. doi:10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. 2012. *Journal of Statistical Software*, 48, 18. doi:10.18637/jss.v048.i04

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent Variable models. *Psychometrika*, 82, 904–927. doi:10.1007/s11336-017-9557-x

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018b). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53, 453–480. doi:10.1080/00273171.2018.1454823

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, 139, 1342. doi:10.1037/a0031808

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. doi:10.1016/j.eurpsy.2018.12.005

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, 126, 969–988. doi:10.1037/abn0000276

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review*, 112, 60–74. doi:10.1037/0033-295X.112.1.60

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31, 271–288. doi:10.1080/1047840X.2020.1853461

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, 12, 999–1020. doi:10.1177/1745691617705892

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., … Karstoft, K.-I. (2018). Replicability and generalizability of posttraumatic stress disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science*, 6, 335–351. doi:10.1177/2167702617745092

Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2014). Depression is more than the sum score of its parts: Individual DSM symptoms have different risk factors. *Psychological Medicine*, 44, 2067–2076. doi:10.1017/S0033291713002900

Fried, E. I., van Borkulo, C. D., & Epskamp, S. (2020). On the importance of estimating parameter uncertainty in network psychometrics: A response to Forbes et al. (2019). *Multivariate Behavioral Research*, 1–6. doi:10.1080/00273171.2020.1746903

Fritz, J., de Graaff, A. M., Caisley, H., van Harmelen, A.-L., & Wilkinson, P. O. (2018). A systematic review of amenable resilience factors that moderate and/or mediate the relationship between childhood adversity and mental health in young people. *Frontiers in Psychiatry*, 9, 230. doi:10.3389/fpsyt.2018.00230

Fritz, J., Stochl, J., Fried, E. I., Goodyer, I. M., van Borkulo, C. D., Wilkinson, P. O., & van Harmelen, A. L. (2019). Unravelling the complex nature of resilience factors and their changes between early and later adolescence. *BMC Medicine*, 17, 203. doi:10.1186/s12916-019-1430-6

Galanaki, E. P., & Vassilopoulou, H. D. (2007). Teachers and children's loneliness: A review of the literature and educational implications. *European Journal of Psychology of Education*, 22, 455. doi:10.1007/BF03173466

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337–1345. doi:10.1097/00004583-200111000-00015

Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, 7, 125–130. doi:10.1007/s007870050057

Goodman, R. J., Samek, D. R., Wilson, S., Iacono, W. G., & McGue, M. (2019). Close relationships and depression: A developmental cascade approach. *Development and Psychopathology*, 31, 1451–1465. doi:10.1017/S0954579418001037

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116. doi:10.1037/a0038889

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, 1–55. doi:10.1080/10705519909540118

Hyde, J. S., Mezulis, A. H., & Abramson, L. Y. (2008). The ABCs of depression: Integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychological Review*, 115, 291–313. doi:10.1037/0033-295X.115.2.291

Johnstone, L. (2018). Psychological formulation as an alternative to psychiatric diagnosis. *Journal of Humanistic Psychology*, 58, 30–46. doi:10.1177/0022167817722230

Jones, P. B. (2013). Adult mental health disorders and their age at onset. *British Journal of Psychiatry*, 202, s5–s10. doi:10.1192/bjp.bp.112.119164

Kalisch, R., Cramer, A. O. J., Binder, H., Fritz, J., Leertouwer, I., Lunansky, G., … van Harmelen, A.-L. (2019). Deconstructing and reconstructing resilience: A dynamic network approach. *Perspectives on Psychological Science*, 14, 765–777. doi:10.1177/1745691619855637

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, 73, 52–62. doi:10.1016/j.intell.2018.12.004

Kaufman, T. M. L., Kretschmer, T., Huitsing, G., & Veenstra, R. (2020). Caught in a vicious cycle? Explaining bidirectional spillover between parent-child relationships and peer victimization. *Development and Psychopathology*, 32, 11–20. doi:10.1017/S0954579418001360

Lereya, S. T., Humphrey, N., Patalay, P., Wolpert, M., Böhnke, J. R., Macdougall, A., & Deighton, J. (2016). The student resilience survey: Psychometric validation and associations with mental health. *Child and Adolescent Psychiatry and Mental Health*, 10, 44. doi:10.1186/s13034-016-0132-5

Lin, S.-Y., Fried, E. I., & Eaton, N. R. (2020). The association of life stress with substance use symptoms: A network analysis and replication. *Journal of Abnormal Psychology*, 129, 204–214. doi:10.1037/abn0000485

Masselink, M., Van Roekel, E., Hankin, B. L., Keijsers, L., Lodder, G. M. A., Vanhalst, J., … Oldehinkel, A. J. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, 32, 653–671. doi:10.1002/per.2179

Masten, A. S., & Barnes, A. J. (2018). Resilience in children: Developmental perspectives. *Children (Basel, Switzerland)*, 5, 98. doi:10.3390/children5070098

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, 22, 491–495. doi:10.1017/S0954579410000222

McElroy, E., Fearon, P., Belsky, J., Fonagy, P., & Patalay, P. (2018). Networks of depression and anxiety symptoms across development. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57, 964–973. doi:10.1016/j.jaac.2018.05.027

McElroy, E., & Patalay, P. (2019). In search of disorders: Internalizing symptom networks in a large clinical sample. *Journal of Child Psychology and Psychiatry*, 60, 897–906. doi:10.1111/jcpp.13044

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, doi:10.3758/s13428-020-01398-0

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568–592. doi:10.1037/0021-9010.93.3.568

Meeus, W. (2016). Adolescent psychosocial development: A review of longitudinal models and research. *Developmental Psychology*, 52, 1969–1993. doi:10.1037/dev0000243

Merikangas, K. R., He, J.-p., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. Adolescents: Results from the national comorbidity survey replication–adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49, 980–989. doi:10.1016/j.jaac.2010.05.017

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology. *This Time Forever. Measurement: Interdisciplinary Research and Perspectives*, 2, 201–218. doi:10.1207/s15366359mea0204_1

NHS Digital. (2018). *Mental Health of Children and Young People in England, 2017 Summary of key findings*. Retrieved from https://files.digital.nhs.uk/F6/A5706C/MHCYP%202017%20Summary.pdf

Oh, Y., Greenberg, M. T., Willoughby, M. T., Vernon-Feagans, L., Greenberg, M. T., Blair, C. B., … The Family Life Project Key, I. (2020). Examining longitudinal associations between externalizing and internalizing behavior problems at within- and between-child levels. *Journal of Abnormal Child Psychology*, 48, 467–480. doi:10.1007/s10802-019-00614-6

Patalay, P., & Fitzsimons, E. (2016). Correlates of mental illness and wellbeing in children: Are they the same? Results from the UK millennium cohort study. *Journal of American Academic Child Adolescent Psychiatry*, 55, 771–783. doi:10.1016/j.jaac.2016.05.019

Patalay, P., & Fitzsimons, E. (2017). *Mental ill-health among children of the new century: Trends across childhood with a focus on age 14. September 2017*. London: Centre for Longitudinal Studies.

Patalay, P., & Fitzsimons, E. (2018). Development and predictors of mental ill-health and wellbeing from childhood to adolescence. *Social Psychiatry and Psychiatric Epidemiology*, 53, 1311–1323. doi:10.1007/s00127-018-1604-0

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *British Journal of Psychiatry*, 207, 15–22. doi:10.1192/bjp.bp.114.149591

Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., … Viner, R. M. (2016). Our future: A lancet commission on adolescent health and wellbeing. *The Lancet*, 387, 2423–2478. doi:10.1016/S0140-6736(16)00579-1

Petrides, K. V., & Furnham, A. (2009). *Technical manual for the trait emotional intelligence questionnaires (TEIQue)*. London: London Psychometric Laboratory.

Rapee, R. M., Oar, E. L., Johnco, C. J., Forbes, M. K., Fardouly, J., Magson, N. R., & Richardson, C. E. (2019). Adolescent development and risk for the onset of social-emotional disorders: A review and conceptual model. *Behaviour Research and Therapy*, 123, 103501. doi:10.1016/j.brat.2019.103501

Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., … European Kidscreen Group. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Review of Pharmacoeconomics & Outcomes Research*, 5, 353–364. doi:10.1586/14737167.5.3.353

Rhemtulla, M., Cramer, A., van Bork, R., & Williams, D. R. (2018). Cross-lagged network models. 1-32.

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50, 353–366. doi:10.1017/S0033291719003404

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1, 27–42. doi:10.1177/2515245917745629

Saint-Georges, Z., & Vaillancourt, T. (2020). The temporal sequence of depressive symptoms, peer victimization, and self-esteem across adolescence: Evidence for an integrated self-perception driven model. *Development and Psychopathology*, 32, 975–984. doi:10.1017/S0954579419000865

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health*, 2, 223–228. doi:10.1016/S2352-4642(18)30022-1

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. doi:10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4, 1208–1214. doi:10.1038/s41562-020-0912-z

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11, 702–712. doi:10.1177/1745691616658637

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): A Rasch analysis using data from the Scottish health education population survey. *Health and Quality of Life Outcomes*, 7, 15. doi:10.1186/1477-7525-7-15

Suldo, S., Thalji, A., & Ferron, J. (2011). Longitudinal academic outcomes predicted by early adolescents' subjective well-being, psychopathology, and mental health status yielded from a dual factor model. *The Journal of Positive Psychology*, 6, 17–30. doi:10.1080/17439760.2010.536774

Thompson, R. A. (2019). Emotion dysregulation: A theme in search of definition. *Development and Psychopathology*, 31, 805–815. doi:10.1017/S0954579419000282

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., … Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111, 87. doi:10.1073/pnas.1312114110

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9, 486–492. doi:10.1080/17405629.2012.686740

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17, 228–243. doi:10.1037/a0027127

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, 2, 214–227. doi:10.1177/2515245919848684

WHO. (1946). *constitution of the world health organization*. Retrieved from Geneva: http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf?ua=1

WHO. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision). https://icd.who.int/browse11/l-m/en

Wolff, J. C., & Ollendick, T. H. (2006). The comorbidity of conduct problems and depression in childhood and adolescence. *Clinical Child and Family Psychology Review*, 9, 201–220. doi:10.1007/s10567-006-0011-3

Wolpert, M., & Rutter, H. (2018). Using flawed, uncertain, proximate and sparse (FUPS) data in the context of complexity: Learning from the case of child mental health. *BMC Medicine*, 16, 82. doi:10.1186/s12916-018-1079-6