Methodological and Developmental Studies into the Bifactor Model of Psychopathology

Matthew Paul Constantinou

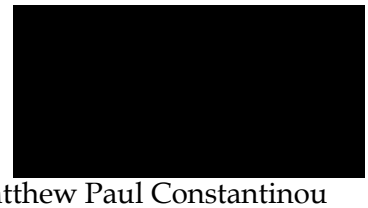
DClinPsy thesis (Volume 1), 2022

University College London

UCL Doctorate in Clinical Psychology

Thesis Declaration Form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: 

Name: Matthew Paul Constantinou

Date: 29/06/22

Overview

This thesis examines the methodological, developmental, and clinical relevance of the bifactor model of psychopathology. The bifactor model organizes mental health problems into a single dimension of psychopathology, e.g., the *p* factor, which captures aspects shared across disorders, and specific dimensions, e.g., internalizing and externalizing, which capture aspects shared among subgroups of disorders.

Part 1 is a reliability meta-analysis of bifactor studies of psychopathology. It uses model-based reliability indices to evaluate how the variance in factor models is distributed and whether this resembles a bifactor structure. Part 2 is a developmental analysis of the relationship between socioeconomic status and the general and specific psychopathology factors. It examines the mediating role of stressful life events, and moderating role of family obligation, in this relationship. Part 3 is a clinically informed evaluation of quantitative models of psychopathology, covering issues related to methodology, epistemology, and clinical application.

Part 1 shows that whilst psychopathology measures tend to be multidimensional (e.g., include both general and specific sources of variance), most of what may be measured in practice reflects the *p* factor. Part 2 shows negative links between socioeconomic status and the *p* factor and specific externalizing factor, and that these links are partially explained by stressful life events, particularly for children who prioritize their families' needs and views. Finally, Part 3 argues for more process-based mental health assessments, idiographic and transactional analyses of symptoms, and a focus on the socio-evolutionary role of communication in assessment and treatment.

Impact Statement

Mental health difficulties such as depression, anxiety, and substance misuse are currently diagnosed using criteria established by consensus. The evidence-base around psychiatric diagnoses has changed dramatically over the last half century and challenges many of the assumptions held about psychiatric diagnoses, including their categorical and discrete nature. This thesis supports an evidence-based shift in conceptualizing and measuring mental health outcomes as hierarchically organized dimensions.

Impact for Mental Health Research.

Research into the aetiology of mental health problems ultimately depends on our 'map' of human suffering. Currently, we divide human suffering into multiple, discrete disorders. Research is therefore focused on discovering biological, psychological, and social mechanisms associated with specific disorders. However, this thesis and the research that it builds on show that psychiatric disorders are more alike than they are different. This thesis demonstrates that self-report measures assess both the shared and unique characteristics of mental health problems (rather than multiple, discrete problems). Accounting for these shared characteristics is important when examining the social determinants of mental health difficulties; associations between socioeconomic status and specific problems may in fact be driven by features shared across mental health problems. The bifactor model is a powerful tool for establishing what is truly unique to specific problem domains and what is the product of shared aetiological factors.

Impact for Clinical Practice

Like mental health research, assessment and treatment is governed by current diagnostic systems that focus on discrete diagnoses. However, this thesis and surrounding research suggest that people rarely report single problems, and that even if they did, there would be widespread differences between two people meeting the criteria for the same disorder. This thesis supports the importance of examining both the shared aspects of mental health problems (e.g., the degree of dysregulation across emotional, cognitive, and behavioural domains) and unique characteristics (e.g., specific problems and characterological tendencies). Evidence-based assessments offer a more precise and comprehensive way of characterising a person's experience, which can improve prognosis by matching treatments more suited to people's needs.

Impact for Mental Health in Society

The findings in this thesis suggest that socioeconomic disadvantage has widespread effects on children's mental health as well as specific effects on behavioural difficulties. This is explained, in part, by an increased incidence of stressful life events. Understanding the pathways towards mental health difficulties can inform public health interventions, which may be most effective at reducing multiple issues through preventing and minimizing the impact of stressful life events. This is more relevant than ever following the Covid-19 pandemic, which has created multiple stressors for families and widened pre-existing inequalities.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank the following people for supporting me personally and professionally throughout the last three years. We have faced one of the greatest challenges of our lifetime. As difficult as it has been, I could not have been luckier to have faced it with them.

First and foremost, I thank my supervisor, Peter Fonagy. It has almost been seven years since you and Liz invited me to discuss my romanticized research ideas. How time flies (and romanticism persists)! I am indebted to you for continuing to support my research career, which was inevitably going to lag during this training. It has been fun as ever to develop novel experimental tasks and psychometric analyses with you; I only regret that we did not have as much time to play due to training demands. You picked me up during the countless times I dropped off the radar and continued to inspire me with your creative ideas and unparalleled insight. It has been an honour to be supervised by you.

I thank those at the Psychoanalysis Unit, including Liz Allison, Chloé Campbell, Patrick Luyten, and Jilly Raw, for inspiring the thought behind this research and their valiant efforts in its development. I also thank my course tutors: Kathy White, for being genuinely curious and compassionate, and Vaughan Bell, for your brief but invaluable support.

I thank my clinical supervisors: Catherine Athanasiadou-Lewis, Margot Kalmus, Louis Acker, Meenaxi Patel, and Emma Goodman. I am particularly grateful to Emma Goodman for challenging me to consider what is important for clinical research and for believing in me to make a difference.

There are several talented clinicians that I have shared this journey with: Adam, Alice, Ally, Amrita, Amy, Andrea, Charlotte, Georgie, Helen, James, Jack G, Jack S, Jordan, Josie, Lara, Leah, Leanne, Maxie, Neil, Rachel, Sam, Shona, Vee, and Zaariyah. I am particularly grateful to Adam (for being a champ), Alice (for having that connection since the interview), Amrita (for your contagious love of life), Amy (for all the relationship chats), Jordan (for being my bestie), Josie (for not actually being boring), Lara (for all the heavenly cake), Rachel (for being *both* super reflective

and watching Love Island), Shona (for supercharging my learning), and Zaariyah (for all the Bryan Adams),

I am also grateful to my dear friends Mattia, Fran, Mila, Jaskiran, and Gabor; Chris, Mossy, and Mike; Alex and Sive; and Jonny, Krissy, Petros, George, Archie, and Demi. You have been there for me throughout the good times and the bad, and I look forward to making many more memories with you.

I want to thank my Mum and Dad, Niki, Mario, and Stephen, and Γιαγια, for your enduring love and support. You mean the world to me. I also want to thank Laur for being so caring and encouraging throughout these last months of write-up. You have shown me a different side to life, one that excites, inspires, and astonishes. It is hard to see a future without you and I want to take our relationship further.

I dedicate this thesis to my Aunty Cindy, whose fun-loving energy and silky red hair will never be forgotten. May you rest in peace.

Part 1. Literature Review

Evaluating Bifactor Models of Psychopathology Using Model-Based Reliability Indices

Note to examiner. This review builds on a previous reliability review of bifactor studies (Constantinou & Fonagy, 2019). The current review advances the former by updating the study list and using mixed-effect models rather than simple averages to pool the reliability indices.

1.1 Abstract

In this chapter, I evaluate methodological questions related to the bifactor model of psychopathology. Bifactor models have become popular over the last decade, but it is uncertain whether psychopathology measures resemble such a model. I ran a meta-analysis of 68 bifactor studies using model-based reliability indices (Explained Common Variance, Omega Hierarchical, Construct Reliability and Factor Determinacy) to evaluate how the variance in item responses is distributed between general and specific psychopathology factors. I found that psychopathology measures tend to be multidimensional, with 61% of the modelled variance explained by the *p* factor and 39% explained by specific factors. Most of the variance in raw total scores (78%) is explained by the *p* factor, while less of the variance in raw subscale scores (39%) is explained by specific factors. Indicators explained 95% of the variance in the *p* factor, and 70% of the variance in specific factors. The *p* factor (r = .95) and specific factors (r = .87) overlapped strongly with observed factors scores. The percentage of uncontaminated correlations, measure composition, indicator, method and sample type, and estimator predicted variation in the general and specific factors' reliability. Overall, psychopathology measures generally resemble bifactor models, but the extent to which we measure general over specific features depends on our scoring and measurement methods.

1.2 Introduction

Over the last thirty years, there has been a growth of research into transdiagnostic models of psychopathology (see Dalgleish, Black, Johnston, & Bevan, 2020, for a review). Transdiagnostic models define broader dimensions that underpin multiple, co-occurring psychiatric diagnoses (Nolen-Hoeksema & Watkins, 2011). For example, half of people who report experiences of depression also report experiences of anxiety and somatization (Löwe et al., 2008). These conditions are thought to be underpinned by common mechanisms that are summarized by an internalizing dimension (Achenbach & McConaughy, 1992; Eaton et al., 2013). Similarly, experiences of antisociality, impulsivity and attentional problems tend to aggregate and can be summarized by an externalizing dimension (Achenbach & McConaughy, 1997; Krueger, Markon, Patrick, Benning, & Kramer, 2007).

Comorbidity does not stop at the level of disorders. Internalizing and externalizing dimensions tend to positively co-occur: people who score highly on internalizing disorders also score highly on externalizing conditions (Lilienfeld, 2003). One reason why internalizing and externalizing dimensions overlap is because they may share a common set of mechanisms summarized by a general psychopathology dimension (Kotov et al., 2017; Lahey et al., 2012). General psychopathology is thought to reflect individual differences in the propensity to develop any and all forms of common mental health problems (Caspi & Moffitt, 2018; Caspi et al., 2014), or an index of impairment (Smith et al., 2020). The mechanisms that underpin a general risk for psychopathology are proposed to be distinct from those that predispose people towards specific internalizing or externalizing conditions (Lahey et al., 2017).

1.2.1 What is the bifactor model?

Bifactor models are a statistical method for decomposing the covariation among observed variables into common and specific latent sources (Holzinger & Swineford, 1937; Reise, 2012). For example, the positive associations among a set of questionnaire items about depression can be summarized by a general factor that reflects individual differences on the broad construct running through all items (e.g., overall depression severity), as well as more specific factors that reflect common themes among groups of items (e.g., low mood, anhedonia, somatic problems). Specific factors are assumed to be orthogonal to the general factor; that is, they capture additional covariation that remains after the variance shared by all variables has been explained by the general factor (Markon, 2019).

The bifactor model was first developed by Holzinger and Swineford (1937) who extended Charles Spearman's two-factor model of intelligence. Spearman hypothesised that a general factor runs through all intelligence test variables, as well as factors unique to each test variable (Spearman's model is equivalent to today's single-factor model; the factors unique to each variable are equivalent to latent error terms). The general factor was dubbed the *g* factor for general intelligence and was thought to reflect a unitary trait that explains individual differences in performance on intelligence tests (Spearman, 1904). Karl Holzinger, Spearman's student, added 'group factors' to Spearman's single-factor model to capture commonalities among sub-groups of intelligence tests (e.g., all tests involving a visuo-spatial component would load on a *g* factor as well as a specific visuo-spatial factor).The bifactor model has been applied to broadband psychopathology measures[1] on the basis that

[1]A broadband measure is any measure that assesses multiple domains of a construct. I use this term to describe psychopathology measures that assess multiple 'spectra' of

various symptoms or disorders form a positive manifold, i.e. they are positively associated with each other (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011, but see van Bork, Epskamp, Rhemtulla, Borsboom, & van der Maas, 2017). Lahey et al. (2012) and Caspi et al. (2014) were first to estimate a general factor that described individual differences in the positive covariation among disorders–the general psychopathology (‘*p*’) factor. They also estimated specific factors that captured the positive covariation among groups of disorders, such as internalizing and externalizing, once the variance common to all disorders was explained by the *p* factor.

A higher-order model can be used instead of a bifactor model to estimate a general psychopathology factor. The higher-order model (also known as the second-order model) was developed by Thurstone (1944), who argued for multiple, correlated domains of intelligence (e.g., visuo-spatial, verbal, working memory, processing speed) that are in turn underpinned by a ‘higher-order’ general factor (e.g., *g* factor). In the context of psychopathology, internalizing and externalizing factors would be understood as products of a higher-order general psychopathology factor.

psychopathology, e.g., internalizing, externalizing, thought disorder, etc. In turn, I use the term ‘spectra’ or ‘spectral domain’ to indicate a measure that goes beyond the disorder level as per Kotov et al.’s (2017) HiTOP model (e.g., the internalizing spectrum captures commonalities among mood and anxiety disorders).

1.2.2 The general factor in the higher-order model is estimated from the correlations among specific factors. By contrast, the general factor in the bifactor model is estimated directly from the observed variables, as are the specific factors. Consequently, the specific factors in the bifactor model are distinct from, and estimated side-by-side with, the general factor, and are what remain after the common variance running through all observed variables is explained by the general factor. Both bifactor and higher-order models are used to estimate the structure of psychopathology and are often treated as competing models. More information about the similarities and differences between these models is described in Part 3 (see 3.2 Quantitative Approaches to Psychopathology).Is there support for a bifactor model of psychopathology?

There are now over 100 studies applying the bifactor model to broadband psychopathology measures. Generally, the bifactor model fits the data better than alternative models across children (Lahey et al., 2015; Moore et al., 2020; Olino et al., 2018), adolescents (Castellanos-Ryan et al., 2016; Laceulle, Vollebergh, & Ormel, 2015; Patalay et al., 2015), and adults (Caspi et al., 2014; Gluschkoff, Jokela, & Rosenström, 2019; Lahey et al., 2012; Lahey et al., 2017). Consistent with theory, the *p* factor is positively associated with a broad set of risk factors for psychopathology (e.g., familial mental health, childhood maltreatment/victimization, and socio-economic status) and these risk factors are no longer associated with specific internalizing and externalizing factors once the *p* factor is accounted for, indicating distinct aetiologies (Caspi et al., 2014; Lahey et al., 2012; Schaefer et al., 2018).

Contrary to the argument that the *p* factor is a statistical artifact, it has been validated against several functional outcomes (Cervin et al., 2021; Pettersson, Lahey, Larsson, & Lichtenstein, 2018; Sallis et al., 2019), shows broadly consistent neural correlates (Moberget et al., 2019; Romer et al., 2018, 2019), heritability estimates of around 50% (Harden et al., 2020; Neumann et al., 2016; Rosenström et al., 2019), and longitudinal stability estimates on par with psychological traits (Greene & Eaton, 2017; Gluschkoff, Jokela, & Rosenström, 2019; McElroy, Belsky, Carragher, Fearon, & Patalay, 2018; Snyder, Young, & Hankin, 2017). Specific factors have also been validated against external measures, and sometimes show stronger predictions than the general factor (Haltigan et al., 2018), but tend to be weaker and less reliable.

Despite the growing support for a bifactor structure of psychopathology, the model is not without issues, including its tendency to overfit noise in the data, higher fitting propensity, difficulties interpreting the general factor, variable *p* factor loadings within and between studies, and weak or vanishing specific factors (Bonifay, Lane, & Reise, 2017; Bornovalova, Choate, Fatimah, Petersen, & Wiernik, 2020; Greene et al., 2019; Sellbom & Tellgen, 2019; Watts, Poore, & Waldman, 2019). It is also uncertain how to interpret the specific psychopathology factors. Most have interpreted specific factors as purer indices of specific problem domains, minus general levels of distress or severity (Patalay et al., 2015; Snyder, Young, & Hankin, 2017). However, some have conceptualized specific factors as psychopathology-less clusters of personality traits (e.g., internalizing as neuroticism and introversion; externalizing as extraversion, low agreeableness, and low conscientiousness; Caspi et al., 2014). Personality traits are not pathological in and of themselves. It is only when they create distress in one's life do they become problematic, which is why it may be helpful to separate out specific, stylistic patterns of relating to general levels

of distress using the specific and general factors, respectively, in the bifactor model (Livesley, 2011).An issue which has yet to be addressed is how well a bifactor structure is measured in practice. Studies applying the bifactor model to psychopathology have so far assumed that our measures reliably capture both general and specific sources of variance. However, our measures may capture general psychopathology at the expense of specific psychopathology dimensions. Or perhaps our measures primarily capture specific psychopathology dimensions whilst the common variance is present but minimal, like background radiation. How do measurement characteristics, such as the number of items, response format, respondent, and measure type, affect the relative strength of the general and specific sources of variance? Model fit indices tell us how well we replicate the correlation matrix with our models, but they do not tell us about *the specific characteristics* of our modelled factors, and whether they resemble a bifactor structure.

1.2.3 Reliability indices: what are they and how do they resolve the issues outlined above?

Model-based reliability indices are a set of measures that can be used to evaluate the internal properties of measures (Rodriguez, Reise, & Haviland, 2016b). They address questions that model fit indices cannot, such as how much of the modelled variance in a measure is attributable to the general vs. specific factors (Explained Common Variance; *ECV*), the extent to which raw total and subscale scores reflect general and specific factors, respectively (Omega Hierarchical; ω_H), how well the general and specific factors variables are represented by their indicators (Construct Reliability; *H*), and how indeterminate or stable observed factor scores are (Factor Determinacy; *FD*). Reliability indices are useful in

determining the extent to which our psychopathology measures resemble a bifactor structure, which includes both general and specific sources of variance.

Rodriguez, Reise, and Haviland (2016a) applied reliability indices to fifty studies that used bifactor models to evaluate the factor structure of psychological, psychopathology and personality measures. The measures assessed broad target constructs (e.g., depression, narcissism) but included several subdomains, and hence were generally assumed to be multidimensional. On average, the general factors accounted for 67% of the modelled variance, suggesting that measurement models mainly reflected the broad target constructs. Furthermore, the variance in raw total scores was largely explained by the general factors, as was the variance in raw subscale scores. In other words, raw subscale scores did not reliably represent the target subdomain beyond the general construct. Finally, the reliability of observed and latent factor scores was high for the general factors on average, but not for the specific factors.

1.2.4 Research Questions

Rodriguez et al.'s (2016a) findings suggest that psychological measures tend to capture the broad target construct they are designed to assess, despite sampling from multiple subdomains. Can we say the same for measures of psychopathology that feature multiple diagnoses or problem domains? I set out to address four questions by applying reliability indices to bifactor studies of broadband psychopathology measures published to date.

Question 1. Do measurement models of broadband psychopathology measures resemble a bifactor structure? Bifactor models may fit the data well, but that does not mean that responses on broadband psychopathology measures

resemble a bifactor structure and hence are equally explained by general and specific factors. If both general and specific sources are important, then we would expect that roughly half of the modelled variance in broadband measures to be attributable to a general *p* factor, and the other half to specific psychopathology factors. If, however, broadband measures are more unidimensional, then we would expect the majority of modelled variance (e.g., ≥ 70%; Rodriguez et al., 2016a) to be explained by a general *p* factor, despite sampling from multiple domains of psychopathology. Alternatively, the majority of variance in psychopathology measures (e.g., ECV ≥ 70%) may be best explained by a set of distinct specific factors, with a relatively weak general *p* factor.

Question 2. Do total and subscale scores reliably reflect variation in the general and specific psychopathology factors, respectively? If general psychopathology is a measurable dimension akin to constructs like general intelligence that run through all items, then we would expect the variance in raw total scores from broadband psychopathology measures to reflect a common source (e.g., $\omega_H \geq .80$; Rodriguez et al., 2016a). Similarly, if specific problem domains can be precisely assessed beyond the common variance, then we would expect raw subscale scores to mainly reflect specific sources (e.g., $\omega_{HS} \geq .80$). Nonetheless, psychopathology subscales might be subject to the same fate as subconstructs in measures of single-domain constructs, their measurement drowned out by a general psychopathology dimension (e.g., Rodriguez et al., 2016a).

Question 3. Can we reliably estimate observed and latent factor scores for the general and specific psychopathology factors? As research into the substantive basis of the *p* factor grows, researchers need to consider how well their measures specify individual differences in the general and specific psychopathology factors,

and how replicable these factors are across studies. This is particularly relevant for specific factors, which are specified by fewer indicators and hence naturally have less information. If general and specific psychopathology factor scores are reliable, then we would expect the majority of variance in latent variables to be explained by their indictors (e.g., $H \geq .70$; Hancock & Mueller, 2001; Rodriguez et al., 2016b).

Researchers may also ask how viable it is to estimate observed factor scores for the general and specific psychopathology factors when it is not possible to specify a bifactor model in a complex measurement model. If the general and specific psychopathology factors are well represented by their observed factor scores, then we would expect the correlation between the factors and factor scores to be high (e.g., $FD \geq .90$; Gorsuch, 1983; Rodriguez et al., 2016a).

Question 4. What methodological characteristics predict variability in the reliability of general and specific psychopathology factors? Testing the latent structure of any psychological domain is dependent on the measures used to estimate it (Marsh & Hau, 2007). The bifactor model of psychopathology is no different. Various methodological characteristics can affect the reliability of latent variables (Achenbach, 2021). In the extreme case, the *p* factor's reliability might simply be a function of the extent to which a measure prioritizes correlations that are solely explained by a general factor. I therefore investigated how methodological characteristics predicted variation in the reliability indices, including the respondent, sample type, sample origin, measure type, measure composition, indicator type, estimator type, percentage of uncontaminated correlations, and publication date.

1.3 Method

1.3.1 Search Strategy

A literature search was conducted in July 2019 (and updated in May 2020, July 2021, and January 2022) using PubMed (n = 651, Jan 2002 – Jan 2022), PsycInfo (including PsycArticles, n = 650, Oct 2002 to Jan 2022), and MEDLINE (n = 650, Oct 2002 – Jan 2022) to identify studies that applied the bifactor model to psychiatric symptoms or disorders. Titles and abstracts were searched using the terms: (bifactor OR bi-factor OR nested factor OR p factor) AND (psychopathology OR psychiatr* OR disorder OR symptom OR diagnosis OR mental). The search produced a total of 1,951 papers published between January 2002 to January 2022, with publications increasing year-upon-year since Caspi et al.'s (2014) seminal paper. Studies were also identified with a citation search of Caspi et al. and Lahey et al.'s (2012) studies using Google Scholar.

Studies were included if they: (i) modelled symptoms or disorders from more than one spectral domain using a bifactor model (e.g., studies that analysed depression, anxiety, and substance problems cover two spectral domains, internalizing and externalizing problems, whereas studies that analysed depression, anxiety and somatic complaints cover a single domain, internalizing), (ii) used confirmatory bifactor analysis, and (iii) provided a standardized factor loading matrix (either in print or via correspondence).

Studies were excluded if they: (i) used exploratory bifactor analysis, exploratory structural equation modelling, or principal components analysis; (ii) used a dataset that was previously analysed with a bifactor model; and (iii) included personality or personality disorder indicators, or solely included well-being

measures. Given that the analysis of broadband psychopathology data with bifactor models is relatively new, studies were not excluded based on the type of estimator used (e.g., maximum likelihood vs. weighted least-squares), or whether their solution included cross-loadings, residual correlations between indicators, or specific factor correlations.

For studies that reported multiple versions of a bifactor model, I chose the model that the authors endorsed (typically based on model fit) and/or was most well specified (based on factor loading patterns). For studies that used overlapping datasets, I included the first study to publish a bifactor model with that dataset and excluded all others to minimize dependencies between estimates. I accounted for dependencies within studies that reported bifactor models for different samples, or longitudinal studies that reported a separate bifactor model at different ages for a given sample, using multilevel models (see Meta-analysis below).

1.3.2 Reliability Coefficients

Explained Common Variance (ECV). *ECV* reflects the proportion of variance modelled among all the factors that is attributable to the general factor (Reise, Moore, & Haviland, 2010). It reflects the 'strength' of the general factor relative to the specific factors or the degree to which a measurement model is unidimensional (Ten Berge & Sočan, 2004). *ECV* ranges from 0-1, with values closer to 1 reflecting more variance explained by the general factor compared to the specific factors (Reise, Scheines, Widaman, & Haviland, 2013). *ECV-subscale* (ECV_S) can be used to compute the proportion of modelled variance explained by a given specific factor. When ECV/ECV_s values are $\geqslant .7$, the modelled variance is said to be 'essentially

unidimensional', i.e. would fit a single-factor model without substantial bias from assessing multiple domains (Rodriguez et al., 2016b).

Coefficient Omega (ω). ω reflects the proportion of variance in raw total scores explained by all factors modelled (McDonald, 1999; Reise, Bonifay, & Haviland, 2013). Omega, like Cronbach's α, is a measure of internal consistency (not unidimensionality) but is calculated from the model-based factor loading matrix rather than the observed variance-covariance matrix (Rodriguez et al., 2016b). Cronbach's α assumes there are equal relationships between the items (i.e. equal factor loadings), whereas ω does not. Omega ranges from 0-1 and increases as the factor loadings and number of items increase.

Omega hierarchical (ω_H). ω_H reflects the proportion of variance in raw total scores explained by the general factor (Reise et al., 2013). ω_H is used an index of general factor saturation, or the extent to which total score variance represents a general source despite a scale's multidimensionality (Reise et al., 2010). In this sense, one can have a multidimensional measurement model (i.e. moderate *ECV* values) but high general factor saturation (i.e. high ω_H values), as total scores can be influenced by a general source over and above specific sources. By dividing ω_H by ω, we can determine the proportion of error-free variance in total scores explained by the general factor (Dueber, 2017; Rodriguez et al., 2016b).

We can also calculate the proportion of variance in raw subscale scores attributable to a specific factor with omega hierarchical subscale (ω_{HS}). ω_{HS} reflects the amount of subscale variance influenced by a specific factor over and above the influence of a general factor. Both ω_H and ω_{HS} range from 0-1, with higher values reflecting a higher proportion of total or subscale variance explained by the general

or specific factor, respectively. Rodriguez et al. (2016a) suggested that ω_H and ω_{HS} values ⩾.8 reflect the dominance of a given factor on scale scores.

Construct Reliability (H). *H* is the proportion of variance in a factor explained by its indicators (Hancock & Mueller, 2001). *H* reflects how well a set of indicators represent a factor, and hence how well that factor would replicate in another study using the same indicators (Rodriguez et al., 2016b). *H* ranges from 0-1, with higher scores indicating that a factor is more well-defined by its indicators. When *H* values are ⩾.7, the factors modelled are thought to be reliably represented by their indicators (Hancock & Mueller, 2001).

Factor Determinacy (FD). Factor scores are indeterminate, i.e. there exists an infinite number of observed factor scores for any set of latent factors (Gutmann, 1955). Factor determinacy (*FD*) reflects the extent to which factor scores are good estimates of individual differences on the factor (Grice, 2001). One way to calculate *FD* is by reproducing the correlation matrix from factor loadings to estimate the correlation between the factor and observed factor scores (Rodriguez et al., 2016b). *FD* ranges from 0-1, with higher values reflecting higher correspondence between the (model-implied) factor scores and their factors. Gorsuch (1983) suggested that *FD* values ⩾ .9 reflect trustworthy factor score estimates.

Percentage of Uncontaminated Correlations (PUC). PUC describes the number of correlations among indicators that can be uniquely explained by a general factor, above and beyond the specific factors (Rodriguez et al., 2016b). It is not so much a reliability index but a structural indicator of the extent that the composition of a measure favours the measurement of a general factor that is 'uncontaminated' by the shared influence of specific sources (i.e.

multidimensionality). PUC ranges from 0-1, with higher values reflecting a higher proportion of unique correlations that are solely explained by the general factor, relative to the correlations explained by the general and specific factors. PUC values ⩾.7 indicate that the correlation matrix is structured towards a general source (Rodriguez et al., 2016a). PUC is largest when there are multiple specific factors, each with a small number of indicators. Provided that the indicators load well onto the general factor, the proportion of modelled variance (*ECV*) and total score variance (ω_H) explained by the general factor is likely to be high when the PUC is high. PUC also moderates the impact of *ECV* on parameter bias when fitting a unidimensional model to multidimensional data (Reise et al., 2013).

1.3.3 Statistical Analysis

Data Aggregation. I ran a meta-analysis of reliability coefficients, also known as a psychometric meta-analysis or reliability generalization study (Vacha-Haase, 1998). First, standardized factor loading matrices were extracted from each study and analysed with Dueber's (2017) bifactor indices calculator, a freely available Excel-based tool for calculating the model-based reliability indices described by Rodriguez et al. (2016a, 2016b). There is also an *R* package available, which I used for factor loading matrices with more than 100 indicators (Dueber, 2021). There may be slight differences between the reliability estimates reported here and those reported in the original papers because I reproduced the correlation matrices from the factor loading matrices rather than from the raw data. There have been a limited number of reliability generalization studies using model-based reliability indices, so analysis decisions were based on studies using coefficient alpha, the closest analogue to model-based reliability indices.

Data Transformation. I transformed the reliability indices due to their strong negative skew which is characteristic of reliability coefficients (e.g., values tend to cluster around values of .70; Greco, O'Boyle, Cockburn, & Yuan, 2018). I used Bonett's (2002; 2010) transformation for coefficient alpha, as it provides more stable confidence intervals compared to Hakstian-Whalen's *T* transformation and is suitable for small samples. The Bonett and *T* transformations tend to perform similarly in simulation studies (López-Pina, Sánchez-Meca, & López-López, 2013), and produce similar estimates to untransformed coefficients if heterogeneity is low (Sánchez-Meca, López-López, & López-Pina, 2013). I ran sensitivity analyses using the *T* transformed and raw reliability coefficients for comparison.

Model Type. I used a random-effects model to pool the reliability indices because psychopathology bifactor studies vary widely in their samples and measures. In random-effects models, each study is assumed to approximate a reliability coefficient from a range of possible populations, rather than a single population (Borenstein, Hedges, Higgins, & Rothstein, 2010). In other words, rather than assuming that the *p* factor has one fixed level of reliability in the population, we assume that the *p* factor's reliability differs depending on the population it is sampled from.

I estimated a three-level random-effects models for *p* factor reliability estimates, with random intercepts at level 1 (i.e. within-study variability due to sampling and measurement error), level 2 (i.e. between-sample variability, as some studies contributed multiple *p* factor estimates from different samples or time-points), and level 3 (i.e. between-study variability due to sampling for multiple populations). Furthermore, I estimated a four-level random-effects model for specific factor reliability estimates, with random intercepts at level 1 (within-study

variability), level 2 (i.e. between-specific factor variability since each study reported multiple specific factor reliabilities), level 3 (i.e. between-sample variability), and level 4 (i.e. between-study variability). For a more detailed description of the models, see Appendix 1.2.1). Analyses were run using the *metafor* package (v 3.4) in *R* (Viechtbauer, 2010).

Weighting Method. I weighted reliability estimates by the inverse of the sampling variance. Studies that provide more reliable (i.e. less noisy) estimates are given more weight when calculating the pooled estimate (Borenstein et al., 2010). When effect sizes are reliability coefficients, sampling variances should also be adjusted for the measurement instruments' properties (e.g., number of items, strength and homogeneity of the inter-item correlations, Rodriguez & Maeda, 2006). The Bonett (and *T*) transformation incorporates these factors into its sampling variance calculations.

Heterogeneity Estimator. I used the Paule and Mandel (PM) method to estimate the between-study variance (τ^2) since it does not rely on distributional assumptions (reliability coefficients tend to be negatively skewed) and tends to be less biased and more efficient compared to other estimators, particularly under higher levels of heterogeneity (Boedeker & Henson, 2020; Langan, Higgins, & Simmonds, 2016; Veroniki et al., 2016). Furthermore, confidence intervals around τ^2 were estimated using Hartung's (1999) weighted variance method, as has been used in prior reliability generalizations of coefficient alpha (Blázquez-Rincón et al., 2021; López-Pina et al., 2015; Rubio-Aparicio et al., 2020)

I tested whether the amount of heterogeneity was substantial using Paule and Mandel's (1982) adapted version of Cochran's *Q* statistic, which captures the

degree of heterogeneity based on the weighted sum of differences between each study's reliability estimate and the mean reliability estimate. Q_{PM} follows a chi-square distribution with k-1 degrees of freedom (df); Q values that exceed a critical value suggest that studies differ substantially in their reliability estimates due to unexplained differences in population estimates (e.g., $\tau^2 > 0$), rather than random error alone (i.e. within-study variability)

I also quantified the degree of heterogeneity using the I^2 statistic, which reflects the percentage of variance in reliability estimates due to between-study variability compared to within-study variability (Higgins, Thompson, Deeks, & Altman, 2003). I^2 values of 25%, 50%, and 75% reflect low, moderate, and high levels of heterogeneity (Higgins et al., 2003).

Meta-Regression. I ran multilevel meta-regressions to determine what methodological characteristics predicted variation in the reliability indices. The meta-regression models were an extension of the three-level and four-level meta-analytic models used to estimate the pooled reliability indices for the p factor and specific factors, respectively (for a more detailed model description, see Appendix B). Models were estimated using Restricted Maximum-Likelihood (REML).

Linearity of relationships was checked with predicted plots; normality and homoscedasticity of residuals was checked with quantile plot sand predicted vs. residual plots, respectively[2]; and collinearity among predictors was checked with the Variance Inflation Factor (VIF; where values > 2.5 warrant concern, Allison, 2012) and Spearman's Rho rank order correlations > | .5| (Dormann et al., 2012).

[2]Residual plots are only available using univariate meta-analyses in the *metafor* package. I therefore estimated the predicted-residual plots and quantile plots using residuals from a univariate model.

Potential outliers were identified using studentized residuals ($t > |3|$; Osbrone & Overbay, 2004) and predicted plots. Influential cases were identified using DFFITS which reflects the amount (in standard deviations) that the pooled effect size changes after removing a given study, and therefore how influential that study is on the pooled effect (Belsley, Kuh, & Welsch, 2005). DFFITS values $> |.31|$ for *p* factor reliabilities and $> |.16|$ for specific factor reliabilities (i.e. $3\sqrt{1/k-1}$), where *k* is the number of entries; Viechtbauer, 2010).

The initial set of predictors included: sample size (log), mean sample age (centred), PUC (centred), number of items per factor (log), publication date (centred), respondent (self vs. other, including caregiver, teacher, clinician and mixed self-and-other reports), indicator type (symptom- vs. subscale-level or symptom count), method type (questionnaire vs. interview), sample type (community vs. clinical, community vs. population), measure construction (single vs. multiple measures), indicator measurement level (binary vs. ordinal/continuous indicator), continent (North America vs. other, including UK, Europe, Australasia, and South America), and estimator type (WLSMV vs. MLR/Bayes, because factor loadings can be inflated under WLSMV when latent distributions are highly skewed; Li, 2016).

To calculate the strength of associations between reliability indices whilst accounting for the multilevel data structure, I regressed each reliability index on one other index in a multilevel meta-regression. This produced a regression weight, which is the equivalent to a correlation coefficient, but also accounts for the nesting among data levels.

There is currently no method for back-transforming regression slopes. I therefore report marginal means for a one-unit increase in the predictor of interest, holding the other predictors constant, as an estimate of a back-transformed slope. Pseudo-R^2 was calculated as the proportional reduction in between-study variability between a multilevel meta-regression model with and without predictors.

1.4 Results

Eighty-one studies met the inclusion criteria (see Figure 1.1). Of these, three studies did not provide standardized factor loading matrices in print or by correspondence; three studies used exploratory bifactor analysis/structural equation modelling; two studies used bifactor growth models with a single indicator per psychopathology domain (aggregating over time would result in a single loading); one used a tri-factor model which included personality and personality disorder indicators in addition to psychopathology indictors; two reported mis-specified models; and one was not published in English. Study characteristics for the remaining 68 studies published between January 2009 and January 2022 are summarized in Table 1.1. A full list of studies can be found in Table 1.2.

The total number of participants included was 1,083,652, 71% of which came from Du Rietz et al.'s (2020) sample (n = 774,416). The average sample size was 11,172 (3,221 excluding Du Rietz et al.) but varied widely between studies (SD = 78,579; SD = 6,611 excluding Du Rietz et al.). Most studies were balanced for sex (49% male; k = 63) but recruited mainly Caucasian samples (72% Caucasian; k = 43). Just over half of the samples originated from North America. The average sample age was 16 (SD = 10), with less than half being children (2-12; 42%), a third being

adolescents (13-17; 33%) and a quarter being adults (18-47; 25%). Most samples were community-based (64%), 23% were recruited from outpatient or inpatient clinics, and 13% were randomly recruited from the population.

Figure 1.1

PRISMA Flow Diagram

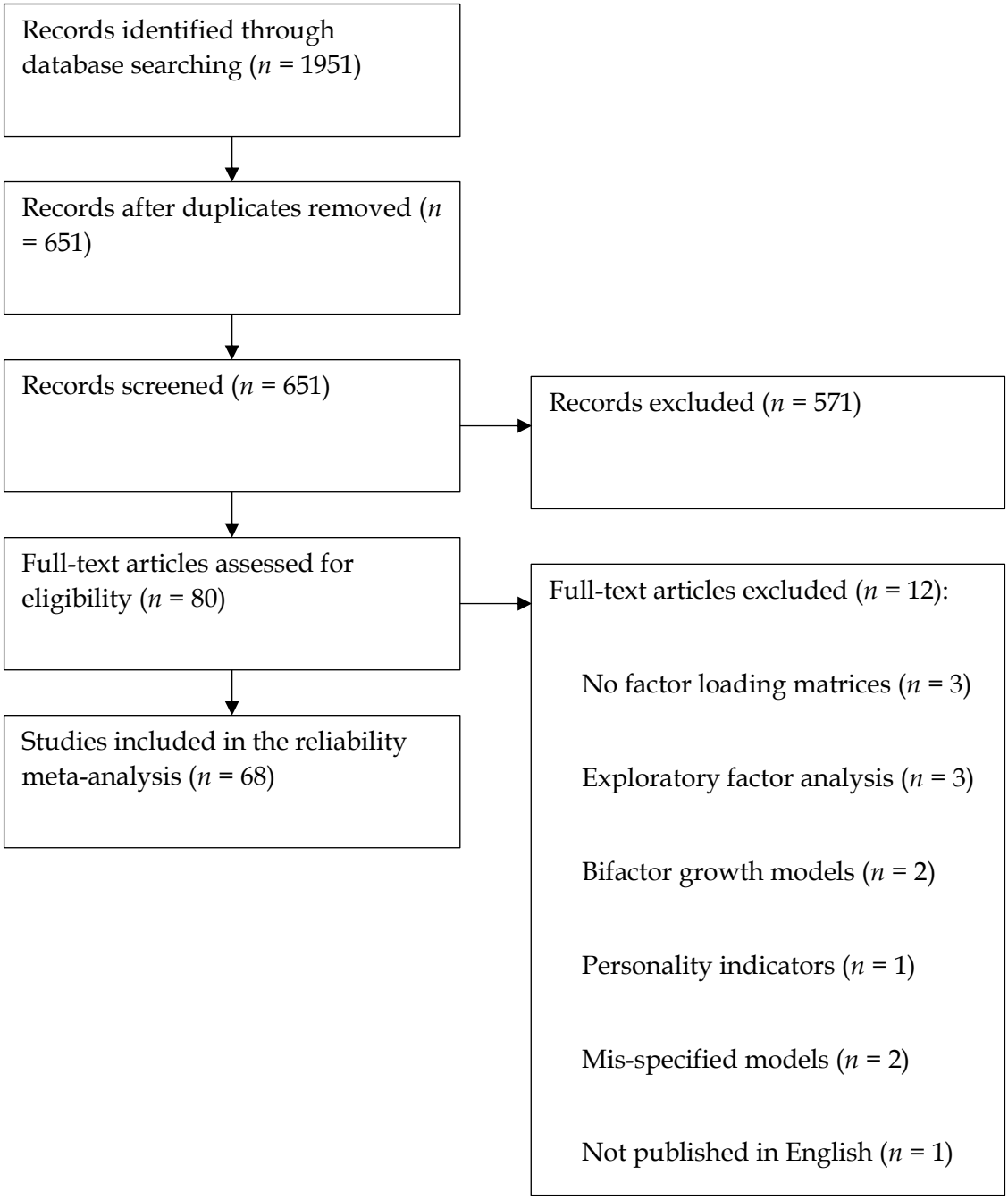


Table 1.1

Methodological Characteristics for the 97 entries belonging to 68 studies

Study Characteristic	*M* or *n*	*SD* or %
Age (years; 2-47)	16	10
Childhood (2-12)	41	42%
Adolescence (13-17)	32	33%
Adulthood (18-47)	24	25%
N (120-774,416)	11,172	78,579
Caucasian	72%	39%
Male (vs. female)	49%	16%
Region		
North America	49	51%
UK	11	11%
Europe	29	30%
Australasia	4	4%
South America	4	4%
Sample Type		
Community	62	64%
Clinical	22	23%
Population	13	13%
Respondent Type		
Self	43	44%
Caregiver	37	38%
Teacher	8	8%
Clinician	1	1%
Multiple	8	8%
Indicator Type		
Item-level	52	54%
Subscale-level	45	46%
Measure Type		
Questionnaire	71	73%
Interview	20	21%
Mixed	5	5%
Medical Records	1	1%
Multiple (vs. single) measures	29	30%
Ordinal (vs. binary) indicators	82	85%
WLSMV (vs. MLR/Bayes)	50	53%
Publication date (mode; 2009-2021)	2017	22%
Number of specific factors (2-15)	3.4	2.3
Number of items (5-139)	32	28
Item-level (12-139)	50	28
Subscale (5-28)	11	4
PUC (0.38-0.92)	0.70	0.14

Note. *M* = mean; MLR = Robust Maximum Likelihood estimator; *n* = sample size; PUC = Percentage of Uncontaminated Correlations; *SD* = standard deviation; WLSMV = Weighted Least Squares Means and Variances estimator.

Most measures were questionnaires (73%). Most respondents were either the participants themselves (44%) or their caregivers (38%). Over half of measures were analyzed at the item/symptom level (54%) compared to the subscale/disorder level (46%), but the majority (85%) of indicators fell on the ordinal rather than binary scale. Seventy percent of studies used a single measure rather than combining multiple measures. The average number of specific factors estimated was 3.4 (*SD* = 2.3).

1.4.1 Explained Common Variance (*ECV*)

Tables A1.1 and A1.2 in Appendix 1 present the pooled reliability indices for the *p* factor and specific factors, respectively, including Bonett-transformed, *T*-transformed, and raw estimates. I report the Bonett-transformed estimates below, which were similar to the *T*-transformed estimates; raw estimates were slightly smaller than the untransformed estimates.

ECV tells us the proportion of variance in modelled item responses explained by the *p* factor compared to the specific factors. If broadband psychopathology measures conform to a bifactor structure with both general and specific sources, then the *p* factor should explain roughly 50% of the variance in item responses. The pooled *ECV* was 0.61 (95% CI = .58-.64). Hence, the *p* factor accounted for 61% of the modelled variance in psychopathology measures on average (leaving 39% of the variance explained by specific factors). There was a substantial amount of heterogeneity (total I^2 = 99%), mainly due to the high amount of between-study heterogeneity (level 3 I^2 = 79%). There was a negligible amount of

sampling error (level 1 variability = .34%) and a small amount of between-sample heterogeneity (level 2 I^2 = 21%).

The pooled *ECVs* was .16 (95% CI = .14-.17); on average, a given specific factor accounted for 16% of the modelled variance in psychopathology measures. Like the *p* factor, there was a substantial amount of heterogeneity (total I^2 = 99%) mainly due to between-study differences (level 4 I^2 = 69%). Unlike the *p* factor, however, the remainder of the heterogeneity was explained by differences between specific factors (level 2 I^2 = 28%) rather than differences between time-points/samples within a study (level 3 I^2 = 0%). *ECVs* for the specific internalizing factor was .19 (95% CI = .16-.21; *k* = 69) and .20 (95% CI = .18-.23, *k* = 65) for the specific externalizing factor.

1.4.2 Omega Coefficients (ω/ω_H)

ω_H values tell us the proportion of variance in raw total scores explained by the *p* factor, whilst ω_{HS} values tell us the proportion of variance in raw subscale scores explained by a given specific factor above and beyond the *p* factor. Relative ω_H and ω_{HS} values reflect the ratio between ω_H/ω_{HS} and ω/ω_S, hence the proportion of error-free total and subscale variance explained by the *p* factor and relevant specific factor, respectively. If variability in total and subscale scores reflects variation in the *p* factor or relevant specific factor, respectively, then ω_H/ω_{HS} will be high (e.g., ≥ .80).

We first need to determine how much of the variance in raw total scores was explained by both the *p* factor and specific factors (i.e. ω/ω_S). The pooled ω was 0.94 (95% CI = .92-.95), hence 94% of the variance in raw total scores could be explained by the general and specific psychopathology factors. In turn, only 6% of the variance in total scores was attributable to error (e.g., item-specific variance). Similarly, 87% of the variance in raw subscale scores was attributable to the general and specific psychopathology factors, leaving 13% of the variance explained by error.

On average, the *p* factor accounted for 78% (95% CI = .74-.81) of the variance in raw total scores with error (ω_H) and 84% (95% CI = .81-.87) of the variance in total scores without error (relative ω_H). Therefore, the *p* factor explained most of the individual variation in total scores across studies. As with *ECV* estimates, heterogeneity was high (total I^2 = .99%), with a negligible amount of within-study variability (level 1 variability = .10%), a small amount of between-sample heterogeneity (level 2 I^2 = 16%), and a substantial amount of between-study heterogeneity (level 3 I^2 = 84%).

On average, a given specific factor accounted for 39% (95% CI = .34-.42) of the raw subscale variance including error (ω_{HS}), and 49% (95% CI = .43-.54) of the subscale variance without error (relative ω_{HS}). Therefore, the *p* factor accounted for at least half of the variability in raw subscale scores. There was substantial heterogeneity in ω_{HS} estimates (total I^2 = 99%), most of which was explained by within-study differences between specific factor estimates (level 2 I^2 = 72%) rather than between-study differences in pooled estimates (level 4 I^2 = 27%). It was not uncommon for studies to describe specific factors that accounted for almost none of the variance in their subscales and others that accounted for at least half (e.g., Hyland et al., 2018; Preti, Carta, & Petretto, 2019; Du Rietz et al., 2020; St Clair et al., 2017; Urban, Arrindell, Demetrovics, Unoka, & Timman, 2016). ω_{HS} for the specific internalizing factor was .36 (95% CI = .29-.41; *k* = 69) and .43 (95% CI = .37-.48, *k* = 65) for the specific externalizing factor.

1.4.3 Construct Reliability (*H*) and Factor Determinacy (*FD*)

H tells us the proportion of variance in a factor predicted by its indicators, whilst *FD* reflects the correlation between a factor and observed factor scores. If factors (and hence latent factor scores) are well represented by their indicators, then *H* values will be high (e.g., ≥ .70). Moreover, if observed factor scores provide reliable estimates of factors, then *FD* values should be high (e.g., ≥ .90).

For *H* values, 92% (95% CI = .90-.94) of the variance in the *p* factor could be explained by its indicators, demonstrating high levels of construct reliability. Heterogeneity in estimates (total I^2 = 99%) was largely explained by differences between studies (level 3 I^2 = 93%). Specific factors were also reliable, with 70% (95% CI = .66-.74) of the variance explained by their indicators. Heterogeneity was

explained both by between-study differences (level 4 I^2 = 59%) and between-specific factor differences (level 2 I^2 = 41%). The specific externalizing factor met the threshold for reliability (*H* =.71, 95% CI = .65-.76; *k* = 65), while the internalizing factor fell just under (*H* = .69, 95% CI = .63-.74; *k* = 69).

A similar picture emerged with factor determinacy, where the *p* factor and its factor scores showed high correspondence (*FD* = .95, 95% CI = .93-.95). Unlike *H* values, heterogeneity in estimates was split between differences in time-points/samples (level 2 I^2 = 45%) and between-study variability (level 3 I^2 = 55%). The reliability of specific factor scores fell just below the suggested cut-off (*FD* = .87, 95% CI = .85-.89). Heterogeneity was also split between studies (level 4 I^2 = 54%) and specific factors (level 2 I^2 = 45%). Neither the specific internalizing factor (*FD* = .86, 95% CI = .82-.89; *j/k* = 65) nor externalizing factor (*FD* =.86, 95% CI = .83-.89; *j/k* = 69) met the suggested cut-off for factor determinacy.

Table 1.2

Study characteristics and model-based reliability estimates for bifactor studies of psychopathology published between January 2009 and January 2022

Author	Method	Sample	Items	Factor	*ECV(s)*	*ω(s)*	ω_H(s)	*Rel. ω*	*H*	*FD*	*PUC*
Urban et al. (2014)	A. SCL-90	D. 2710	83	Global distress	0.84	0.99	0.97	0.98	0.99	0.99	0.89
	B. Questionnaire	E. 40	12	Somatic	0.06	0.95	0.37	0.39	0.80	0.93	
	C. Item	F. Self	10	OC	0.01	0.92	0.03	0.04	0.27	0.64	
		G. Population	9	IS	0.01	0.91	0.05	0.05	0.26	0.65	
			13	Depression	0.02	0.94	0.09	0.10	0.53	0.82	
			10	Anxiety	0.02	0.94	0.08	0.09	0.48	0.81	
			6	Hostility	0.02	0.90	0.16	0.18	0.56	0.87	
			7	Phobia	0.02	0.93	0.17	0.18	0.56	0.89	
			6	Paranoia	0.01	0.87	0.09	0.11	0.33	0.69	
			10	Psychoticism	0.01	0.93	0.02	0.02	0.25	0.65	
Miller et al. (2019)	A. CBCL, ADOS	D. 415	32	Dysreg profile	0.80	0.97	0.93	0.96	0.97	0.98	0.58
	B. Questionnaire	E. 3	8	Anx/depressed	0.09	0.89	0.33	0.37	0.70	0.87	
	C. Item	F. Caregiver	19	Aggressive	0.06	0.97	0.02	0.02	0.58	0.88	
		G. Community	5	Attention	0.06	0.84	0.31	0.38	0.60	0.88	
Sheldrick et al. (2012)	A. PPSC	D. 646	18	*p*	0.80	0.96	0.91	0.94	0.96	0.97	0.84
	B. Questionnaire	E. 3	4	Internalizing	0.08	0.90	0.21	0.23	0.53	0.81	
	C. Item	F. Caregiver	6	Externalizing	0.07	0.89	0.26	0.29	0.53	0.82	
		G. Clinical	3	Attention	0.06	0.92	0.25	0.27	0.47	0.83	
		D. 165	9	*p*	0.79	0.88	0.81	0.92	0.93	0.97	0.56

Tang et al. (2020)	A. ASR, LSAS, BAI, BDI B. Questionnaire C. Subscale	E. 26 F. Self G. Community	5 4	Internalizing Externalizing	0.13 0.08	0.88 0.68	0.09 0.16	0.10 0.23	0.47 0.31	0.82 0.58	
Geeraerts et al. (2015)	A. CBCL B. Questionnaire C. Item	D. 247 E. 5 F. Caregiver G. Clinical	32 8 19 5	Dysreg. profile Anx/depressed Aggressive Attention	0.76 0.10 0.09 0.05	0.98 0.91 0.98 0.90	0.90 0.41 0.09 0.24	0.92 0.46 0.10 0.27	0.98 0.79 0.74 0.61	0.99 0.94 0.92 0.94	0.58
Laceulle, Vollebergh, & Ormel (2015)*	A. YSR, RCADS, CAPE B. Questionnaire C. Item	D. 2230 E. 15 (11-19) F. Self G. Community	12 6 3	*p* Internalizing Externalizing	0.76 0.10 0.14	0.98 0.95 0.94	0.89 0.17 0.51	0.91 0.18 0.55	0.97 0.55 0.79	0.98 0.89 0.96	0.73
Neumann et al. (2016)	A. CBCL, SRS, CPRS, TRF B. Questionnaire C. Subscale	D. 1954 E. 7 F. Caregiver G. Population	28 10 11	*p* Internalizing Externalizing	0.76 0.07 0.16	0.81 0.55 0.69	0.73 0.06 0.30	0.90 0.11 0.43	0.85 0.32 0.50	0.91 0.57 0.70	0.74
Preti, Carta, & Petretto (2019) - Urban	A. SCL-90 B. Questionnaire C. Item	D. 817 E. 18 F. Self G. Community	83 12 10 9 13 10 6	Global distress Somatic OC IS Depression Anxiety Hostility	0.76 0.02 0.03 0.04 0.04 0.01 0.04	0.97 0.84 0.81 0.86 0.89 0.86 0.83	0.95 0.07 0.16 0.27 0.15 0.02 0.32	0.97 0.09 0.20 0.31 0.16 0.02 0.38	0.97 0.36 0.47 0.61 0.56 0.24 0.63	0.98 0.63 0.73 0.83 0.80 0.58 0.83	0.89

			7	Phobia	0.03	0.72	0.27	0.38	0.49	0.74	
			6	Paranoia	0.02	0.78	0.20	0.25	0.42	0.71	
			10	Psychoticism	0.02	0.80	0.17	0.21	0.46	0.72	
St Clair et al. (2017)	A. Multiple	D. 2228	106	*p*	0.76	0.97	0.92	0.94	0.99	0.99	0.87
	B. Questionnaire	E. 19	13	Self-confidence	0.04	0.93	0.31	0.33	0.72	0.90	
	C. Item	F. Self	9	Antisocial	0.06	0.90	0.56	0.62	0.83	0.93	
		G. Community	7	Worry	0.02	0.94	0.18	0.19	0.55	0.89	
			17	Aberrant thoughts	0.07	0.91	0.48	0.53	0.85	0.93	
			30	Mood	0.06	0.79	0.00	0.01	0.79	0.91	
Urban, Arrindell, Demetrovics, Unoka, & Timman (2016) - Hungary	A. SCL-90	D. 972	83	Global distress	0.75	0.99	0.96	0.97	0.98	0.99	0.89
	B. Questionnaire	E. 35	12	Somatic	0.05	0.94	0.33	0.36	0.76	0.91	
	C. Item	F. Self	10	OC	0.02	0.91	0.06	0.06	0.52	0.81	
		G. Clinical	9	IS	0.03	0.91	0.26	0.29	0.63	0.86	
			13	Depression	0.01	0.94	0.02	0.02	0.35	0.75	
			10	Anxiety	0.02	0.92	0.02	0.02	0.46	0.81	
			6	Hostility	0.04	0.90	0.45	0.50	0.75	0.91	
			7	Phobia	0.04	0.92	0.34	0.37	0.76	0.94	
			6	Paranoia	0.02	0.83	0.27	0.32	0.53	0.81	
			10	Psychoticism	0.02	0.85	0.09	0.10	0.46	0.75	
Preti, Carta, & Petretto (2019) – Rural	A. SCL-90	D. 507	83	Global distress	0.74	0.97	0.94	0.97	0.97	0.98	0.89
	B. Questionnaire	E. 17	12	Somatic	0.06	0.85	0.39	0.46	0.69	0.84	
	C. Item	F. Self	10	OC	0.02	0.80	0.10	0.13	0.38	0.67	
		G. Community	9	IS	0.03	0.84	0.18	0.22	0.53	0.79	

			13	Depression	0.03	0.89	0.10	0.11	0.54	0.77	
			10	Anxiety	0.03	0.84	0.17	0.20	0.52	0.76	
			6	Hostility	0.04	0.76	0.38	0.50	0.59	0.80	
			7	Phobia	0.02	0.75	0.11	0.15	0.40	0.68	
			6	Paranoia	0.02	0.76	0.15	0.19	0.37	0.65	
			10	Psychoticism	0.01	0.81	0.01	0.01	0.31	0.61	
O'Reilly et al. (2020)	A. A-TAC	D. 30444	43	*p*	0.73	0.97	0.89	0.92	0.96	0.96	0.77
	B. Questionnaire	E. 10.5 (9/12)	12	Anxiety/Emotion	0.07	0.86	0.29	0.34	0.65	0.81	
	C. Item	F. Caregiver	10	Opposition	0.04	0.88	0.15	0.17	0.45	0.68	
		G. Population	10	Impulsivity	0.05	0.92	0.18	0.19	0.55	0.74	
			11	Inattention	0.11	0.95	0.30	0.32	0.75	0.86	
Patterson et al. (2021)	A. Multiple	D. 315	26	*p*	0.73	0.97	0.93	0.96	0.96	0.98	0.85
	B. Mixed	E. 14	5	Anxiety	0.02	0.85	0.06	0.07	0.25	0.62	
	C. Item	F. Self	5	Depression	0.03	0.90	0.10	0.12	0.36	0.73	
		G. Community	6	Affective Lability	0.06	0.88	0.14	0.15	0.61	0.87	
			2	Sleep	0.04	0.81	0.44	0.54	0.54	0.82	
			4	Psychosis	0.09	0.93	0.44	0.48	0.73	0.93	
			4	Basic Symptoms	0.03	0.90	0.16	0.18	0.40	0.78	
Cervin et al. (2021)	A. CBCL	D. 480	8	*p*	0.72	0.86	0.79	0.91	0.84	0.92	0.57
	B. Questionnaire	E. 11	4	Internalizing	0.08	0.77	0.01	0.01	0.28	0.68	
	C. Subscale	F. Caregiver	4	Externalizing	0.20	0.78	0.23	0.30	0.65	0.85	
		G. Clinical									

Hankin et al. (2017)	A. CBCL B. Questionnaire C. Subscale	D. 554 E. 8 F. Caregiver G. Clinical	8 3 3	*p* Internalizing Externalizing	0.72 0.15 0.13	0.94 0.85 0.89	0.82 0.33 0.29	0.88 0.39 0.33	0.91 0.58 0.48	0.94 0.87 0.77	0.79
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL B. Questionnaire C. Item	D. 1253 E. 11 F. Caregiver G. Community	68 31 30 7	*p* Internalizing Externalizing Attention	0.71 0.15 0.10 0.04	0.97 0.93 0.96 0.86	0.86 0.37 0.11 0.26	0.89 0.40 0.12 0.30	0.97 0.84 0.78 0.62	0.98 0.92 0.90 0.85	0.60
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL B. Questionnaire C. Item	D. 1253 E. 14 F. Caregiver G. Community	70 31 31 8	*p* Internalizing Externalizing Attention	0.71 0.15 0.10 0.04	0.97 0.93 0.96 0.86	0.86 0.37 0.11 0.26	0.89 0.40 0.12 0.30	0.97 0.84 0.78 0.62	0.98 0.92 0.90 0.85	0.60
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL B. Questionnaire C. Item	D. 1253 E. 5 F. Caregiver G. Community	67 30 29 8	*p* Internalizing Externalizing Attention	0.71 0.15 0.10 0.05	0.96 0.91 0.94 0.83	0.86 0.36 0.09 0.21	0.89 0.39 0.10 0.25	0.96 0.81 0.82 0.65	0.97 0.89 0.94 0.90	0.61
Weissman et al. (2019)	A. CDI, SCARED, YSR, CBCL,PTSD-RI B .Questionnaire C. Subscale	D. 259 E. 12 F. Self + Caregiver	6 3 3	*p* Internalizing Externalizing	0.71 0.21 0.08	0.90 0.82 0.85	0.77 0.31 0.11	0.86 0.38 0.13	0.86 0.61 0.27	0.90 0.82 0.60	0.60

		G. Community									
Deutz et al. (2018)	A. SDQ	D. 768	15	Dysreg. profile	0.70	0.94	0.85	0.90	0.92	0.95	0.71
	B. Questionnaire	E. 14	5	Emotional prob.	0.11	0.82	0.32	0.40	0.55	0.78	
	C. Item	F. Caregiver	5	Conduct prob.	0.08	0.88	0.13	0.14	0.52	0.82	
		G. Community	5	Hyp-inattention	0.11	0.90	0.21	0.23	0.68	0.95	
Moore et al. (2020)	A. CBCL	D. 5934	66	*p*	0.70	0.98	0.87	0.89	0.98	0.98	0.64
	B. Questionnaire	E. 9	32	Internalizing	0.11	0.94	0.42	0.45	0.85	0.92	
	C. Item	F. Caregiver	15	Externalizing	0.13	0.97	0.19	0.19	0.86	0.93	
		G. Population (weighted)	19	Attention	0.06	0.95	0.22	0.23	0.73	0.90	
Calkins et al. (2015)	A. GOASSESS	D. 9498	15	*p*	0.69	0.90	0.81	0.90	0.88	0.93	0.70
	B. Interview	E. 14	5	Anxious-misery	0.05	0.78	0.08	0.10	0.27	0.57	
	C. Subscale	F. Multiple	6	Fear	0.06	0.77	0.08	0.10	0.34	0.63	
		G. Community	4	Behavior prob.	0.20	0.82	0.49	0.60	0.69	0.86	
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL	D. 1253	66	*p*	0.69	0.97	0.85	0.88	0.97	0.98	0.61
	B. Questionnaire	E. 10	31	Internalizing	0.16	0.92	0.39	0.42	0.84	0.91	
	C. Item	F. Caregiver	27	Externalizing	0.10	0.95	0.10	0.11	0.76	0.90	
		G. Community	8	Attention	0.05	0.87	0.27	0.32	0.62	0.85	
Tackett et al. (2013)	A. CAPS	D. 1569	11	*p*	0.69	0.92	0.77	0.84	0.95	0.98	0.71
	B. Interview	E. 14	5	Internalizing	0.20	0.79	0.43	0.55	0.66	0.84	
	C. Subscale	F. Multiple	4	Externalizing	0.12	0.85	0.26	0.30	0.50	0.82	
		G. Community									

Thompson, Richards, Ploubidis, Fonagy, & Patalay (2021) - BCS	A. Rutter A Scale B. Questionnaire C. Item	D. 15258 E. 16 F. Caregiver G. Population	16 5 11	*p* Internalizing Externalizing	0.69 0.15 0.17	0.91 0.70 0.92	0.78 0.44 0.13	0.85 0.62 0.14	0.90 0.66 0.68	0.93 0.83 0.84	0.46
Hyland et al. (2020)	A. ITQ, PHQ-9, GAD-7, SCID-II, AUDIT, APMS, APLSS, ACE, B. Questionnaire C. Item	D. 1051 E. 47 F. Self G. Community	49 13 24 5 7	*p* Fear Distress Externalizing Thought Disorder	0.68 0.07 0.08 0.06 0.10	0.99 0.98 0.98 0.78 0.96	0.91 0.11 0.13 0.70 0.66	0.92 0.11 0.14 0.91 0.69	0.99 0.81 0.79 0.93 0.90	0.99 0.96 0.96 0.98 0.98	0.67
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL B. Questionnaire C. Item	D. 1253 E. 9 F. Caregiver G. Community	65 31 26 8	*p* Internalizing Externalizing Attention	0.68 0.16 0.10 0.06	0.97 0.93 0.95 0.86	0.84 0.35 0.15 0.34	0.87 0.37 0.16 0.39	0.96 0.83 0.76 0.66	0.97 0.90 0.87 0.84	0.61
Morales et al. (2021)	A. CBCL B. Questionnaire C. Subscale	D. 291 E. 9 F. Caregiver G. Community	6 3 3	*p* Internalizing Externalizing	0.68 0.15 0.18	0.71 0.41 0.76	0.56 0.22 0.18	0.78 0.54 0.24	0.83 0.28 0.32	0.90 0.54 0.52	0.6
	A. C-DISC-IV B. Interview	D. 1004 E. 13	10 7	*p* Internalizing	0.68 0.20	0.89 0.85	0.73 0.23	0.82 0.27	0.88 0.56	0.93 0.76	0.47

Oro, Goldsmith, & Lemery-Chalfant (2021)	C. Subscale	F. Self G. Population	3	Externalizing	0.12	0.79	0.28	0.35	0.45	0.72	
Thompson, Richards, Ploubidis, Fonagy, & Patalay (2021) - NCDS	A. Rutter A Scale B. Questionnaire C. Item	D. 17415 E. 16 F. Caregiver G. Population	16 5 11	*p* Internalizing Externalizing	0.68 0.16 0.16	0.91 0.70 0.92	0.76 0.51 0.14	0.84 0.72 0.15	0.90 0.67 0.65	0.93 0.83 0.82	0.46
Wade, Fox, Zeanah, & Nelson (2018)	A. MHBQ B. Questionnaire C. Subscale	D. 220 E. 16 F. Caregiver G. Clinical	8 3 5	*p* Internalizing Externalizing	0.68 0.16 0.16	0.95 0.87 0.95	0.78 0.41 0.21	0.82 0.47 0.22	0.91 0.61 0.62	0.93 0.82 0.94	0.54
Arrindell et al. (2017)	A. SCL-90 B. Questionnaire C. Item	D. 2593 E. 37 F. Self G. Clinical	83 12 10 9 13 10 6 7 6 10	Global distress Somatic OC IS Depression Anxiety Hostility Phobia Paranoia Psychoticism	0.67 0.07 0.03 0.03 0.01 0.04 0.05 0.04 0.02 0.03	0.99 0.92 0.91 0.91 0.93 0.94 0.91 0.92 0.85 0.89	0.94 0.48 0.23 0.24 0.03 0.23 0.55 0.40 0.31 0.25	0.95 0.53 0.25 0.26 0.03 0.24 0.61 0.43 0.37 0.28	0.98 0.82 0.66 0.65 0.43 0.69 0.81 0.73 0.58 0.67	0.99 0.93 0.86 0.88 0.79 0.89 0.94 0.91 0.83 0.86	0.89

Romani-Sponchiado et al. (2021)	A. SDQ	D. 2010	15	*p*	0.67	0.92	0.81	0.89	0.91	0.95	0.71
	B. Questionnaire	E. 12	5	Emotional	0.15	0.78	0.42	0.54	0.62	0.80	
	C. Item	F. Caregiver	5	Conduct	0.12	0.83	0.28	0.34	0.58	0.80	
		G. Community	5	Inattention	0.06	0.87	0.00	0.00	0.35	0.77	
Lahey et al. (2012)	A. AUDADIS-IV	D. 43093	11	*p*	0.66	0.93	0.77	0.83	0.91	0.94	0.71
	B. Interview	E. 35	3	Distress	0.04	0.90	0.10	0.11	0.25	0.58	
	C. Subscale	F. Self	3	Fear	0.06	0.78	0.16	0.21	0.32	0.62	
		G. Population	5	Externalizing	0.24	0.87	0.49	0.56	0.72	0.87	
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL	D. 1253	66	*p*	0.66	0.96	0.82	0.85	0.96	0.97	0.61
	B. Questionnaire	E. 6	31	Internalizing	0.19	0.91	0.48	0.53	0.84	0.91	
	C. Item	F. Caregiver	27	Externalizing	0.09	0.94	0.08	0.08	0.70	0.86	
		G. Community	8	Attention	0.06	0.85	0.32	0.37	0.64	0.83	
Caspi et al. (2014)*	A. DIS	D. 1037	11	*p*	0.65	0.95	0.77	0.81	0.96	0.98	0.76
	B. Interview	E. 27 (18-38)	3	Internalizing	0.07	0.91	0.22	0.24	0.41	0.84	
	C. Subscale	F. Self	5	Externalizing	0.28	0.91	0.59	0.65	0.82	0.95	
		G. Population									
Liu, Mustanski, Dick, Bolland, & Kertes (2017)	A. YSR	D. 592	12	*p*	0.65	0.92	0.77	0.83	0.90	0.94	0.55
	B. Questionnaire	E. 16	6	Internalizing	0.11	0.88	0.05	0.06	0.65	0.94	
	C. Subscale	F. Self	6	Externalizing	0.23	0.87	0.44	0.50	0.69	0.85	
		G. Community									
	A. CBCL	D. 1253	66	*p*	0.64	0.97	0.81	0.84	0.96	0.97	0.61

McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	B. Questionnaire	E. 8	31	Internalizing	0.19	0.92	0.47	0.51	0.86	0.92	
	C. Item	F. Caregiver	27	Externalizing	0.11	0.95	0.14	0.15	0.79	0.91	
		G. Community	8	Attention	0.05	0.86	0.32	0.37	0.65	0.84	
Haltigan et al. (2018)	A. CBCL	D. 2934	78	*p*	0.63	0.98	0.83	0.85	0.97	0.97	0.69
	B. Questionnaire	E. 13	30	Internalizing	0.12	0.95	0.26	0.28	0.83	0.90	
	C. Item	F. Caregiver	29	Externalizing	0.19	0.95	0.48	0.51	0.90	0.94	
		G. Clinical	12	Thought problems	0.04	0.90	0.10	0.11	0.66	0.87	
			7	Attention	0.02	0.82	0.14	0.17	0.54	0.84	
Levin-Aspenson, Watson, Clarke, & Zimmerman (2020) – NCS	A. CIDI	D. 8098	11	*p*	0.63	0.90	0.78	0.86	0.88	0.93	0.84
	B. Interview	E. 33	3	Distress	0.03	0.76	0.06	0.08	0.17	0.49	
	C. Subscale	F. Self	3	Fear	0.12	0.79	0.33	0.41	0.50	0.75	
		G. Population	3	Externalizing	0.23	0.81	0.62	0.76	0.77	0.91	
Mann, Atherton, DeYoung, Krueger, & Robins (2021)	A. DISC-IV	D. 646	9	*p*	0.63	0.88	0.72	0.82	0.86	0.92	0.5
	B. Interview	E. 17	3	Internalizing	0.18	0.75	0.33	0.44	0.85	0.97	
	C. Subscale	F. Self	6	Externalizing	0.19	0.84	0.20	0.23	0.56	0.77	
		G. Community									
	A. FHS	D. 8012	11	*p*	0.63	0.91	0.79	0.87	0.88	0.93	0.68

Martel et al. (2017) – Mothers	B. Questionnaire	E. 36	3	Internalizing	0.09	0.83	0.09	0.11	0.47	0.76	
	C. Subscale	F. Self	5	Externalizing	0.21	0.82	0.53	0.64	0.70	0.85	
		G. Community	3	Thought disorder	0.07	0.78	0.02	0.02	0.45	0.82	
Du Rietz et al. (2020)	A. ICD codes	D. 774416	12	*p*	0.63	0.93	0.79	0.85	0.92	0.96	0.71
	B. Medical Records	E. Missing	5	Internalizing	0.05	0.90	0.02	0.02	0.28	0.69	
	C. Subscale	F. Clinician	8	Externalizing	0.09	0.81	0.30	0.37	0.48	0.81	
		G. Population	4	Neurodev.	0.23	0.86	0.63	0.73	0.79	0.91	
Etkin, Mezquita, Lopez-Fernandez, Ortet, & Ibanez (2020)	A. SENA	D. 835	10	*p*	0.62	0.91	0.72	0.78	0.86	0.91	0.71
	B. Questionnaire	E. 14	5	Internalizing	0.27	0.88	0.47	0.53	0.70	0.84	
	C. Subscale	F. Self	3	Externalizing	0.12	0.81	0.31	0.38	0.48	0.73	
		G. Community									
Hyland et al. (2018)	A. MCMI	D. 420	9	*p*	0.62	0.93	0.81	0.87	0.95	0.99	0.72
	B. Questionnaire	E. 36	4	Internalizing	0.10	0.96	0.04	0.04	0.47	0.95	
	C. Subscale	F. Self	2	Externalizing	0.16	0.74	0.64	0.87	0.67	0.84	
		G. Clinical	3	Thought disorder	0.13	0.80	0.38	0.47	0.63	0.86	
McElroy, Belsky, Carragher,	A. CBCL	D. 1253	60	*p*	0.61	0.96	0.75	0.78	0.95	0.95	0.54
	B. Questionnaire	E. 3	36	Internalizing	0.24	0.93	0.43	0.46	0.87	0.91	
	C. Item	F. Caregiver	19	Externalizing	0.11	0.93	0.23	0.25	0.75	0.84	
		G. Community	5	Attention	0.04	0.79	0.19	0.24	0.63	0.88	

Fearon, & Patalay (2017)											
Vosberg et al. (2021)	A. DISC B. Interview C. Item	D. 1016 E. 15 F. Self G. Community	79 26 20	*p* Internalizing Externalizing	0.62 0.16 0.23	0.97 0.92 0.94	0.83 0.48 0.65	0.85 0.52 0.69	0.97 0.87 0.93	0.98 0.93 0.97	0.83
Brodbeck et al. (2014)	A. BSI B. Questionnaire C. Item	D. 1024 E. 40 F. Self G. Clinical	53 8 10 3 4 3 7 6 12	Global distress Depression Phobia Aggression Suicidal ideation Nervous tension Somatic Info. processing IS	0.61 0.05 0.09 0.04 0.03 0.02 0.06 0.04 0.06	0.98 0.26 0.44 0.55 0.36 0.37 0.46 0.29 0.32	0.90 0.23 0.42 0.42 0.28 0.29 0.42 0.26 0.28	0.92 0.25 0.46 0.52 0.32 0.35 0.48 0.29 0.30	0.97 0.67 0.81 0.68 0.58 0.49 0.70 0.58 0.71	0.97 0.90 0.92 0.91 0.86 0.81 0.88 0.85 0.86	0.86
Wade, Fox, Zeanah, & Nelson (2018)	A. MHBQ B. Questionnaire C. Subscale	D. 220 E. 8 F. Caregiver G. Clinical	8 3 5	*p* Internalizing Externalizing	0.61 0.23 0.16	0.95 0.85 0.96	0.73 0.62 0.20	0.77 0.73 0.20	0.98 0.83 0.57	0.99 0.93 0.93	0.54
Stochl et al. (2015) - ROOTS	A. MFQ, PLIKS-Q, DISC-IV, SCAN 2	D. 1074 E. 17 F. Self	25 13 12	*p* Dep/anxiety Psychotic exp.	0.61 0.27 0.12	0.95 0.94 0.92	0.71 0.39 0.24	0.75 0.41 0.26	0.93 0.84 0.98	0.96 0.91 0.99	0.52

	B. Questionnaire C. Item	G. Community									
Rytilä-Manninen et al. (2016)	A. SCL-90	D. 201	90	Global distress	0.60	0.99	0.92	0.93	0.99	0.99	0.91
	B. Questionnaire	E. 15	12	Somatic	0.07	0.92	0.53	0.58	0.84	0.94	
	C. Item	F. Self	10	OC	0.05	0.94	0.43	0.46	0.80	0.93	
		G. Clinical	9	IS	0.04	0.92	0.36	0.39	0.75	0.92	
			13	Depression	0.03	0.96	0.19	0.20	0.66	0.88	
			10	Anxiety	0.05	0.94	0.40	0.43	0.79	0.93	
			6	Hostility	0.04	0.89	0.49	0.55	0.77	0.92	
			7	Phobia	0.06	0.92	0.61	0.66	0.86	0.95	
			6	Paranoia	0.03	0.85	0.46	0.54	0.69	0.89	
			10	Psychoticism	0.04	0.91	0.36	0.39	0.73	0.89	
McElroy, Belsky, Carragher, Fearon, & Patalay (2017)	A. CBCL	D. 1253	60	*p*	0.59	0.96	0.74	0.77	0.95	0.96	0.54
	B. Questionnaire	E. 2	36	Internalizing	0.28	0.94	0.47	0.50	0.90	0.94	
	C. Item	F. Caregiver	19	Externalizing	0.09	0.93	0.18	0.20	0.72	0.84	
		G. Community	5	Attention	0.04	0.76	0.18	0.24	0.59	0.84	
Aitken et al. (2020)	A. MFQ, RCMAS, LOI-S, ABC	D. 465	37	*p*	0.58	0.97	0.85	0.87	0.97	0.97	0.81
		E. 15	8	Melancholic	0.04	0.93	0.13	0.14	0.51	0.77	
		F. Self	5	Depressive cog.	0.05	0.96	0.27	0.28	0.66	0.93	
	B. Questionnaire	G. Clinical	10	Anxiety	0.07	0.92	0.25	0.28	0.68	0.85	
	C. Item		7	OC	0.14	0.93	0.65	0.70	0.86	0.95	
			7	Conduct prob.	0.13	0.91	0.65	0.71	0.85	0.93	

Deutz et al. (2018)	A. SDQ	D. 768	15	Dysreg. profile	0.58	0.93	0.76	0.82	0.91	0.95	0.71
	B. Questionnaire	E. 7	5	Emotional prob.	0.20	0.82	0.59	0.71	0.73	0.87	
	C. Item	F. Caregiver	5	Conduct prob.	0.13	0.84	0.34	0.41	0.62	0.83	
		G. Community	5	Hyp-inattention	0.08	0.90	0.07	0.07	0.55	0.91	
Deutz et al. (2018)	A. SDQ	D. 768	15	Dysreg. profile	0.57	0.92	0.74	0.80	0.89	0.91	0.71
	B. Questionnaire	E. 10	5	Emotional prob.	0.17	0.79	0.49	0.62	0.68	0.83	
	C. Item	F. Caregiver	5	Conduct prob.	0.11	0.88	0.20	0.23	0.60	0.85	
		G. Community	5	Hyp-inattention	0.16	0.87	0.39	0.44	0.65	0.81	
Schaefer et al. (2018)	A. DIS + others	D. 2066	11	*p*	0.57	0.88	0.70	0.79	0.84	0.89	0.69
	B. Interview	E. 18	5	Internalizing	0.26	0.79	0.49	0.62	0.71	0.85	
	C. Subscale	F. Self	4	Externalizing	0.07	0.76	0.13	0.17	0.31	0.59	
		G. Population	2	Thought disorder	0.10	0.84	0.30	0.35	0.48	0.78	
Martel et al. (2017) - Children	A. DAWBA	D. 2512	15	*p*	0.56	0.92	0.65	0.70	0.88	0.89	0.45
	B. Interview	E. 10	11	Internalizing	0.29	0.89	0.40	0.45	0.75	0.83	
	C. Subscale	F. Caregiver	3	Externalizing	0.15	0.89	0.43	0.48	0.69	0.89	
		G. Community									
Wade, Plamondon, & Jenkins (2021)	A. OCHS, BITSEA	D. 501	6	*p*	0.56	0.52	0.38	0.73	0.52	0.71	0.73
	B. Questionnaire	E. 3	2	Internalizing	0.21	0.26	0.25	0.96	0.28	0.52	
	C. Subscale	F. Caregiver	3	Externalizing	0.23	0.50	0.17	0.34	0.28	0.52	
		G. Community									

Black, Panayiotou, & Humphrey (2019)	A. M&MS, CORS B. Questionnaire C. Item	D. 1982 E. 11 F. Self G. Community	19 9 6 4	*p* Internalizing Externalizing Wellbeing	0.55 0.12 0.20 0.13	0.92 0.87 0.87 0.76	0.74 0.21 0.50 0.44	0.80 0.24 0.57 0.58	0.88 0.58 0.74 0.81	0.88 0.70 0.85 0.97	0.67
Harden et al. (2019)	A. CBCL, CPRS, BFI B. Questionnaire C. Subscale	D. 1913 E. 13 F. Self G. Community	10 3 5 3	*p* Internalizing Externalizing Attention	0.55 0.15 0.20 0.10	0.92 0.80 0.88 0.88	0.70 0.43 0.37 0.29	0.76 0.54 0.43 0.33	0.85 0.62 0.63 0.45	0.90 0.82 0.85 0.79	0.64
Urban, Arrindell, Demetrovics, Unoka, & Timman (2016) - Netherlands	A. SCL-90 B. Questionnaire C. Item	D. 1902 E. 30 F. Self G. Clinical	83 12 10 9 13 10 6 7 6 10	Global distress Somatic OC IS Depression Anxiety Hostility Phobia Paranoia Psychoticism	0.55 0.09 0.05 0.04 0.02 0.05 0.06 0.07 0.03 0.04	0.98 0.89 0.88 0.90 0.93 0.91 0.90 0.91 0.83 0.85	0.90 0.65 0.32 0.27 0.04 0.38 0.65 0.60 0.42 0.28	0.92 0.74 0.36 0.30 0.05 0.42 0.72 0.66 0.51 0.33	0.97 0.84 0.72 0.67 0.54 0.75 0.85 0.83 0.65 0.72	0.98 0.92 0.87 0.88 0.84 0.89 0.94 0.93 0.85 0.87	0.89
Weissman et al. (2020)	A. CDI-2, SCARED, YSR, CBCL B. Questionnaire C. Subscale	D. 120 E. 13 F. Self + Caregiver G. Community	5 2 3	*p* Internalizing Externalizing	0.55 0.22 0.22	0.90 0.97 0.79	0.66 0.40 0.33	0.74 0.41 0.42	0.89 0.72 0.65	0.97 0.93 0.91	0.6

Funkhouser et al. (2021)	A. SCID-I B. Interview C. Subscale	D. 504 E. 22 F. Self G. Community	10 7 3	*p* Internalizing Externalizing	0.53 0.19 0.28	0.75 0.69 0.65	0.54 0.18 0.49	0.73 0.26 0.76	0.71 0.46 0.61	0.83 0.65 0.78	0.47
Snyder, Young, & Hankin (2017)	A. CDI, MASC, CBCL, EAT-QR, SNAP-IV B. Questionnaire C. Subscale	D. 571 E. 14.5 (14-15) F. Multiple G. Community	9 4 3	*p* Internalizing Externalizing	0.53 0.29 0.19	0.92 0.84 0.91	0.70 0.62 0.42	0.76 0.74 0.46	0.85 0.77 0.62	0.90 0.89 0.82	0.75
Gomez, Stavropoulos, Vance, & Griffiths (2019)	A. ADISC-IV B. Interview C. Subscale	D. 866 E. >12 F. Caregiver G. Clinical	13 10 3	*p* Internalizing Externalizing	0.52 0.21 0.27	0.87 0.86 0.81	0.68 0.14 0.75	0.78 0.16 0.93	0.85 0.62 0.84	0.93 0.83 0.92	0.38
Snyder, Young, & Hankin (2017)	A. CDI, MASC, CBCL, EAT-QR, SNAP-IV B. Questionnaire C. Subscale	D. 571 E. 14 F. Multiple G. Community	9 4 3	*p* Internalizing Externalizing	0.52 0.25 0.23	0.90 0.81 0.92	0.69 0.56 0.47	0.76 0.69 0.51	0.84 0.69 0.68	0.90 0.84 0.85	0.75
Wade, Fox, Zeanah, & Nelson (2018)	A. MHBQ B. Questionnaire C. Subscale	D. 220 E. 12 F. Caregiver G. Clinical	8 3 5	*p* Internalizing Externalizing	0.52 0.12 0.35	0.96 0.90 0.96	0.63 0.31 0.53	0.66 0.34 0.55	0.94 0.56 0.83	0.98 0.76 0.96	0.54

Ignatyev, Baggio, & Mundt (2018)	A. MINI, SCID-II B. Interview C. Subscale	D. 427 E. 21 F. Self G. Clinical	10 6 4	*p* Internalizing Externalizing	0.51 0.29 0.20	0.78 0.67 0.74	0.52 0.42 0.24	0.67 0.62 0.32	0.80 0.58 0.58	0.88 0.76 0.77	0.53
Lahey et al. (2015)*	A. CSI, SCARED B. Questionnaire C. Subscale	D. 2450 E. 8.5 (5-11) F. Caregiver G. Community	10 6 5	*p* Internalizing Externalizing	0.51 0.21 0.28	0.89 0.86 0.86	0.60 0.29 0.45	0.67 0.34 0.53	0.82 0.63 0.73	0.85 0.76 0.84	0.44
Constantinou et al. (2019)*	A. SDQ, MFQ B. Questionnaire C. Item	D. 683 E. 15 (14-16) F. Self G. Clinical	20 5 8 6 5	*p* Mood Anxiety Antisocial Attention	0.50 0.13 0.14 0.10 0.12	0.91 0.83 0.78 0.74 0.82	0.79 0.43 0.01 0.20 0.35	0.87 0.52 0.01 0.27 0.43	0.87 0.64 0.63 0.57 0.66	0.92 0.81 0.83 0.77 0.84	0.67
Castellanos-Ryan et al. (2016)	A. DAWBA B. Interview C. Subscale	D. 2144 E. 16 F. Multiple G. Community	12 6 6	*p* Internalizing Externalizing	0.48 0.30 0.21	0.75 0.63 0.74	0.51 0.54 0.05	0.69 0.87 0.07	0.75 0.63 0.52	0.88 0.80 0.78	0.55
Stochl et al. (2015) - ALSPAC	A. MFQ, PLIKS-Q, DISC-IV, SCAN 2	D. 6617 E. 13 F. Self	25 13 12	*p* Dep/anxiety Psychotic exp.	0.48 0.24 0.28	0.95 0.92 0.93	0.64 0.38 0.53	0.67 0.42 0.57	0.89 0.82 0.85	0.84 0.84 0.86	0.52

	B. Questionnaire C. Item	G. Community									
Castellanos-Ryan et al. (2016)	A. DAWBA	D. 2144	12	*p*	0.47	0.81	0.57	0.70	0.75	0.87	0.55
	B. Interview	E. 14	6	Internalizing	0.31	0.72	0.54	0.74	0.66	0.81	
	C. Subscale	F. Multiple	6	Externalizing	0.22	0.75	0.10	0.13	0.60	0.81	
		G. Community									
Shields et al. (2019)	A. CBCL, C-DISC	D. 895	11	*p*	0.47	0.85	0.56	0.66	0.82	0.90	0.55
		E. 12	6	Internalizing	0.22	0.79	0.25	0.31	0.94	0.99	
	B. Interview + Questionnaire	F. Caregiver	5	Externalizing	0.31	0.81	0.57	0.70	0.72	0.86	
	C. Subscale	F. Community									
King et al. (2020)	A. SCID, SOGS-RA	D. 1329	11	*p*	0.44	0.70	0.40	0.57	0.67	0.80	0.55
		E. 25	6	Internalizing	0.12	0.64	0.12	0.19	0.30	0.52	
	B. Interview + Questionnaire	F. Self	5	Externalizing	0.44	0.66	0.61	0.92	0.68	0.82	
	C. Subscale	G. Community									
Olino et al. (2018)	A. PAPA	D. 545	9	*p*	0.44	0.79	0.49	0.62	0.81	0.90	0.56
	B. Interview	E. 3	5	Internalizing	0.21	0.69	0.35	0.51	0.56	0.74	
	C. Subscale	F. Caregiver	4	Externalizing	0.35	0.78	0.55	0.71	0.70	0.85	
		G. Community									
Afzali, Sunderland,	A. SDQ, BSI	D. 3826	36	*p*	0.42	0.95	0.64	0.67	0.91	0.90	0.61
	B. Questionnaire	E. 13	20	Internalizing	0.29	0.94	0.47	0.49	0.90	0.91	

Carragher, & Conrod (2017)	C. Item	F. Self	7	Externalizing	0.08	0.82	0.33	0.40	0.75	0.87	
		G. Community	9	Thought disorder	0.21	0.92	0.69	0.75	0.87	0.92	
Brandes, Herzhoff, Smack, & Tackett (2019)	A. CBCL, C-DISC	D. 695	11	*p*	0.42	0.81	0.52	0.65	0.69	0.79	0.71
	B. Interview + Questionnaire	E. 10	4	Internalizing	0.21	0.67	0.26	0.39	0.81	0.95	
	C. Subscale	F. Caregiver	5	Externalizing	0.38	0.80	0.60	0.75	0.72	0.83	
		G. Community									
Carragher et al. (2016)	A. SDQ, BSI, RAPI, DISC	D. 2175	44	*p*	0.42	0.97	0.70	0.72	0.94	0.96	0.65
	B. Questionnaire	E. 13	20	Internalizing	0.23	0.96	0.44	0.45	0.93	0.96	
	C. Item	F. Self	15	Externalizing	0.19	0.94	0.49	0.52	0.96	0.98	
		G. Community	9	Thought disorder	0.15	0.94	0.66	0.70	0.89	0.95	
Niarchou et al. (2017)	A. K-SADS	D. 331	60	*p*	0.42	0.98	0.71	0.72	0.97	0.98	0.75
	B. Interview	E. 17	17	Mood	0.15	0.96	0.54	0.56	0.93	0.97	
	C. Item	F. Self	9	Anxiety	0.10	0.94	0.60	0.64	0.94	0.98	
		G. Clinical	15	Psychosis	0.11	0.96	0.39	0.41	0.90	0.97	
			19	ADHD	0.23	0.97	0.69	0.71	0.97	0.98	
Patalay et al. (2015)	A. SDQ, M&MS	D. 23447	25	*p*	0.42	0.94	0.57	0.61	0.88	0.88	0.51
	B. Questionnaire	E. 12	14	Internalizing	0.24	0.91	0.42	0.46	0.82	0.87	
	C. Item	F. Self	11	Externalizing	0.34	0.92	0.66	0.73	0.88	0.92	
		G. Community									

Pezzoli, Antfolk, & Santtila (2017)	A. Multiple	D. 13024	9	*p*	0.42	0.86	0.58	0.67	0.77	0.85	0.72
	B. Questionnaire	E. 35	2	Internalizing	0.10	0.82	0.30	0.36	0.40	0.65	
	C. Subscale	F. Self	4	Externalizing	0.29	0.76	0.58	0.77	0.84	0.92	
		G. Population	3	Body	0.19	0.77	0.37	0.48	0.76	0.92	
Gomez, Stavropoulos, Vance, & Griffiths (2019)	A. ADISC-IV	D. 1233	13	*p*	0.41	0.88	0.47	0.54	0.87	0.92	0.38
	B. Interview	E. <12	10	Internalizing	0.35	0.88	0.40	0.46	0.76	0.87	
	C. Subscale	F. Caregiver	3	Externalizing	0.24	0.78	0.73	0.93	0.79	0.91	
		G. Clinical									
Conway, Mansolf, & Reise (2019)	A. Diagnostic Screener	D. 25002	15	*p*	0.41	0.87	0.58	0.67	0.78	0.81	0.57
	B. Questionnaire	E. 22	9	Internalizing	0.24	0.81	0.40	0.49	0.69	0.79	
	C. Subscale	F. Self	4	Externalizing	0.15	0.70	0.40	0.56	0.63	0.79	
		G. Community	3	Eating pathology	0.20	0.88	0.54	0.62	0.73	0.87	
Olino et al. (2018)	A. PAPA	D. 545	9	*p*	0.41	0.79	0.45	0.58	0.75	0.86	0.56
	B. Questionnaire	E. 6	5	Internalizing	0.31	0.75	0.52	0.69	0.65	0.80	
	C. Subscale	F. Caregiver	4	Externalizing	0.28	0.73	0.47	0.64	0.65	0.81	
		G. Community									
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.40	0.95	0.70	0.73	0.93	1.00	0.69
	B. Questionnaire	E. 10	8	Internalizing	0.16	0.85	0.83	0.97	0.87	0.94	
	C. Item	F. Teacher	9	ADHD	0.10	0.97	0.32	0.33	0.75	1.00	

		G. Community	18	Aggression	0.20	0.92	0.54	0.58	0.84	0.94	
			8	Prosociality	0.14	0.84	0.79	0.95	0.83	0.92	
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.36	0.95	0.67	0.71	0.89	0.85	0.69
	B. Questionnaire	E. 12	8	Internalizing	0.16	0.88	0.73	0.84	0.87	0.93	
	C. Item	F. Teacher	9	ADHD	0.15	0.91	0.52	0.57	0.83	0.88	
		G. Community	18	Aggression	0.17	0.92	0.44	0.48	0.81	0.81	
			8	Prosociality	0.15	0.86	0.74	0.87	0.84	0.91	
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.36	0.95	0.66	0.70	0.90	0.86	0.69
	B. Questionnaire	E. 9	8	Internalizing	0.16	0.86	0.81	0.95	0.88	0.94	
	C. Item	F. Teacher	9	ADHD	0.12	0.92	0.47	0.51	0.78	0.85	
		G. Community	18	Aggression	0.19	0.93	0.44	0.47	0.83	0.83	
			8	Prosociality	0.16	0.88	0.74	0.84	0.86	0.92	
Levin-Aspenson, Watson, Clarke, & Zimmerman (2020) - MIDAS	A. SCID	D. 2900	11	*P*	0.35	0.81	0.51	0.62	0.86	0.94	0.65
	B. Interview	E. 39	6	Internalizing	0.17	0.73	0.23	0.31	0.53	0.77	
	C. Subscale	F. Self	3	Externalizing	0.29	0.79	0.70	0.89	0.77	0.89	
		G. Clinical	2	Thought Disorder	0.19	0.75	0.63	0.85	0.67	0.84	
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.35	0.95	0.67	0.70	0.90	0.87	0.69
	B. Questionnaire	E. 7	8	Internalizing	0.15	0.85	0.80	0.94	0.87	0.94	
	C. Item	F. Teacher	9	ADHD	0.14	0.93	0.49	0.53	0.83	0.88	
		G. Community	18	Aggression	0.20	0.93	0.50	0.54	0.84	0.86	
			8	Prosociality	0.16	0.89	0.74	0.82	0.86	0.92	

Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.35	0.95	0.65	0.69	0.89	0.85	0.69
	B. Questionnaire	E. 11	8	Internalizing	0.17	0.87	0.77	0.88	0.86	0.93	
	C. Item	F. Teacher	9	ADHD	0.12	0.91	0.43	0.47	0.78	0.83	
		G. Community	18	Aggression	0.21	0.92	0.50	0.55	0.84	0.85	
			8	Prosociality	0.15	0.85	0.78	0.92	0.84	0.91	
Romer et al. (2017)	A. MINI + others	D. 1246	13	*p*	0.34	0.87	0.50	0.57	0.74	0.82	0.74
	B. Interview	E. 20	5	Internalizing	0.34	0.87	0.60	0.69	0.85	0.92	
	C. Subscale	F. Self	5	Externalizing	0.32	0.83	0.66	0.80	0.81	0.90	
		G. Community									
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.33	0.95	0.64	0.67	0.89	0.84	0.69
	B. Questionnaire	E. 8	8	Internalizing	0.16	0.85	0.84	0.98	0.88	0.94	
	C. Item	F. Teacher	9	ADHD	0.14	0.93	0.49	0.53	0.84	0.88	
		G. Community	18	Aggression	0.19	0.93	0.44	0.47	0.83	0.82	
			8	Prosociality	0.17	0.88	0.82	0.93	0.87	0.93	
Murray, Eisner, & Ribeaud (2016)	A. SBQ	D. 1572	40	*p*	0.32	0.94	0.66	0.70	0.89	0.86	0.69
	B. Questionnaire	E. 13	8	Internalizing	0.18	0.87	0.87	1.00	0.90	0.95	
	C. Item	F. Teacher	9	ADHD	0.13	0.93	0.46	0.49	0.79	0.83	
		G. Community	18	Aggression	0.19	0.92	0.54	0.59	0.85	0.87	
			8	Prosociality	0.18	0.89	0.82	0.92	0.88	0.94	
	A. SBQ	D. 1572	40	*p*	0.31	0.94	0.62	0.66	0.88	0.84	0.69

Murray, Eisner, & Ribeaud (2016)	B. Questionnaire	E. 15	8	Internalizing	0.19	0.87	0.86	0.99	0.89	0.94	
	C. Item	F. Teacher	9	ADHD	0.12	0.92	0.41	0.45	0.76	0.79	
		G. Community	18	Aggression	0.21	0.92	0.55	0.59	0.84	0.86	
			8	Prosociality	0.18	0.86	0.85	1.00	0.86	0.93	
Gibbons, Rush, & Immekus (2009)	A. PDSQ	D. 3791	139	Overall MI	0.30	0.98	0.83	0.84	0.97	0.97	0.92
	B. Questionnaire	E. 41	26	Depression	0.09	0.93	0.66	0.71	0.94	0.97	
	C. Item	F. Self	7	Dysphoria	0.04	0.92	0.71	0.77	0.88	0.95	
		G. Clinical	6	Gen. anxiety	0.02	0.87	0.39	0.44	0.69	0.87	
			11	Agoraphobia	0.05	0.95	0.61	0.64	0.89	0.95	
			14	Panic	0.06	0.96	0.49	0.51	0.90	0.95	
			15	Social anxiety	0.07	0.96	0.58	0.61	0.91	0.96	
			7	PTSD	0.04	0.92	0.67	0.73	0.88	0.95	
			8	OCD	0.03	0.91	0.49	0.54	0.80	0.91	
			5	Somatic	0.02	0.81	0.58	0.71	0.75	0.88	
			5	HYPO	0.03	0.94	0.65	0.69	0.87	0.96	
			7	Alcohol	0.06	0.95	0.91	0.96	0.96	0.98	
			6	Drug	0.06	0.97	0.90	0.93	0.96	0.98	
			10	Bulimia	0.08	0.96	0.84	0.88	0.96	0.98	
			6	Mania	0.03	0.88	0.74	0.85	0.88	0.94	
			6	Psychosis	0.02	0.86	0.44	0.51	0.68	0.85	
Watts, Poore, & Waldman (2019)	A. ECRS	D. 2498	15	*p*	0.30	0.82	0.38	0.47	0.73	0.82	0.63
	B. Questionnaire	E. 9	8	Fear	0.29	0.72	0.64	0.89	0.72	0.84	

C. Subscale	F. Caregiver	2	Distress	0.13	0.80	0.44	0.55	0.67	0.84
	G. Community	5	Externalizing	0.29	0.83	0.53	0.64	0.80	0.88

Note. Studies have been ordered by *ECV* values from highest to lowest. *ECV(s)* = Explained Common Variance (subscale); *FD* = Factor Determinacy; *H* = Construct Reliability; $\omega_{(s)}$ = Coefficient Omega (subscale); $\omega_{H(s)}$ = Coefficient Omega Hierarchical (subscale); PUC = Percentage of Uncontaminated Correlations; Rel. ω = Relative Omega.

*Models collapsed over multiple time-points.

Key: A = assessment measure; B = method (Questionnaire vs. Interview); C = indicator type (Item vs. Subscale); D = sample size; E = average sample age (years); F = respondent (Self; Caregiver; Teacher; Multiple); G = sample type (Clinical, Community, Population).Dep/anxiety = mixed depression and anxiety; Dysreg. = Dysregulation profile; IS = Interpersonal sensitivity; MI = mental illness; neurodev. = neurodevelopmental; prob. = problems;

Measures: ABC = Antisocial Behavior Checklist (DSM-IV); ADISC-IV = Anxiety Disorders Interview Schedule for Children for the DSM-IV; ADOS = Autism Diagnostic Observation Scale; AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule–DSM–IV Version; APMS = Adult Psychiatric Morbidity Survey; APSS = Adult Psychotic-like Sympton Screener; ASR = Adult Self Report; ATAC = Autism–Tics, AD/HD, and Other Comorbidities; BITSEA = Brief Infant–Toddler Social and Emotional Assessment; BFI = Big Five Inventory; BSI = Brief Symptom Inventory; CAPE = Community Assessment of Psychic Experiences; CAPS = Child and Adolescent Psychopathology Scale; CBCL = Child Behavior Checklist; CDI = Children's Depression Inventory; CORS = Child Outcome Rating Scale; CDSC = Computerized Diagnostic Schedule for Children; CPRS = Conner's Parent Rating Scale; CSBQ = Child Social Behaviour Questionnaire; CSI = Child Symptom Inventory; DIS = Diagnostic Interview Schedule; DAWBA = Development and Well-being Assessment; DISC-IV = Diagnostic Interview Schedule for Children-IV; EATQ-R = Aggression scale of the Early Adolescent Temperament Questionnaire-Revised; ECRS = Emory Combined Rating Scale; FHS = Family History Screen; GOASSESS = National Institute of Mental Health Grand Opportunity Assessment; ITQ = International Trauma Questionnaire; K-SADS = Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age; LOIS = Leyton Obsessional Inventory-Short; M&MS = Me & My School Questionnaire; MASC = Manifest Anxiety Scale for Children; MCMI = Millon Clinical Multiaxial Inventory-III; MHBQ = MacArthur Health and Behavior Questionnaire; MINI = Mini-International Neuropsychiatric Interview; PAPA = Preschool Age Psychiatric Assessment; PLIKS-Q = Psychosis-Like Symptom Questionnaire; OCHS = Ontario Child Health Study Scales; PPSC = Preschool Pediatric Symptom Checklist; RAPI = Rutgers Alcohol Problem Index; RCADS = Revised Child Anxiety and Depression Scale; RCMAS = Revised Children's Manifest Anxiety Scale; SBQ = Social Behaviour Questionnaire; SCAN = Schedules for Clinical Assessment in Neuropsychiatry; SCARED = Screen for Anxiety-Related Emotional Disorders; SCID-II = Structured Clinical Interview Axis II for DSM-IV; SENA = Sistema de Evaluación de Niños y Adolescentes; SDQ = Strengths and Difficulties Questionnaire; SNAP-IV = Swanson,

Nolan and Pelham Questionnaire for DSM-IV; SOGS-RA = South Oaks Gambling Screen-Revised for Adolescents; SRS = Social Responsiveness Scale; TRF = Teacher's Rating Form; SCL-90 = Symptom Checklist-90; YSR = Youth Self-Report.

1.4.4 Meta-Regressions

Diagnostics. Relationships between continuous predictors and reliability indices were generally linear. Residuals were approximately normally distributed, which was largely due to the Bonett transformation. There were minimal signs of heteroscedasticity, except for *ECVs* and ω_{HS}, which showed an increase in error with an increase in reliability values. However, this is likely to be offset by the large-sample approximation of the sampling variance, which does not assume homoscedasticity.

Reliability indices were strongly positively correlated with each other (mean r_p = .69, SD_p = .19; mean $r_{specific}$ = .58, $SD_{specific}$ = .16; see Appendix 1, Table A1.3). Overlap between predictors was variable (mean *r* = -.03, *SD* = .23; see Appendix 1, Table A1.4), with most predictors showing a weak relationship, and a minority showing strong relationships (range = 0 to .85). Collinearity was high for the log number of items (VIF = 4.49), indicator type (e.g., item- vs. subscale level; VIF = 4.88), log sample size (VIF = 2.61), and age (VIF = 2.42).

The log number of items was strongly negatively correlated with indicator type (*r* = -.85), as subscale-level indicators naturally include fewer items. Moreover, indicator type was moderately negatively correlated with the method used (e.g., questionnaire vs. interview; *r* = -.54), as interviews were less likely to use item-level indicators. I removed both the number of items and indicator type and kept the method type variable, as the latter subsumes information from both the former and is qualitatively richer.

The log sample size was moderately correlated with the sample type (e.g., community vs. population *r* = .48), as population samples were typically larger on

average. I therefore removed sample size because the sample type variable captures variations in size whilst offering a richer picture of the samples' characteristics.

Finally, age was strongly correlated with the respondent (self vs. other; r = -.69), since caregiver and teacher respondents were over-represented in younger samples. Again, I removed age as it lacks the qualitative richness of respondent type.

The final predictors included PUC (centered), publication date (log), respondent (self vs. other), method (questionnaire vs. interview) sample type (community vs. clinical; community vs. population), measure composition (single vs. multiple measures), indicator type (binary vs. ordinal), sample origin (North America vs. other), and estimator (WLSMV vs. MLR/Bayes).

Predicting variability in p factor reliabilities. Regression coefficients for the three-level meta-regression models predicting *p* factor reliabilities from method variables can be found in Appendix 1, Table A1.5 (level 1 = within-study variance; level 2 = between-sample variance; level 3 = between-study variance). All models significantly fit the data (QMs[10] = 37.02-99.34, ps < .001) and explained 21% of the variance in *ECV* values (pseudo-R^2 level 2 = .18, level 3 = .32), 55% of the variance in ω_H values (pseudo-R^2 level 2 = .58, level 3 = .32), 41% of the variance in *H* values (pseudo-R^2 level 2 = .44, level 3 = -.01), and 31% of the variance in *FD* values (pseudo-R^2 level 2 = .54, level 3 = .03).

PUC values (i.e. the percentage of inter-item correlations uniquely explained by a general factor) consistently but weakly predicted all *p* factor reliability indices, with a 10% increase in PUC values predicting a .07%-2.2% increase in reliability. Ordinal indicators significantly predicted higher *ECV* (+8%), ω_H (+8%), and *H*

values (+4%) compared to binary indicators. Combining multiple measures significantly predicted lower ω_H (-10%), *H* (-6%) and *FD* values (-4%). Using questionnaires rather than interviews significantly predicted higher *ECV* (+7%) and ω_H values (+9%). Using the MLR or Bayes estimator rather than the WLSMV estimator predicted lower ω_H (-14%) and *H* values (-7%). Finally, population-based samples significantly predicted higher *ECV* values compared to community (+10%) and clinical samples (+13%), while clinical samples trended to predict higher (+3%) *FD* values compared to community samples.

Predicting variability in specific factor reliabilities. Regression coefficients for the four-level meta-regression models predicting specific factor reliabilities can be found in Appendix 1, Table A1.6 (level 1 = within-study variance; level 2 = between specific-factor variance; level 3 = between-sample variance; level 4 = between-study variance). Only the *ECVs* model was significant (QM[10] = 138.41, $p < .001$), and explained 57% of the variance in *ECVs* values (pseudo-R^2 level 2 = 4%; level 3 = 33%; level 4 = 80%). The ω_{HS} model was marginally significant (QM[10] = 18.25, $p = .051$) and explained 6% of the variance across levels (pseudo-R^2 level 2 = 0%, level 3 = 44%, level 4 = 23%). Neither the *H* model (QM[10] = 12.87, $p = .231$) nor the *FD* model (QM[10] = 11.89, $p = .292$) were significant, and each explained 2% of the variance in reliability estimates (*H* pseudo-R^2 level 2 = 0.4%, level 3 = 4%, level 4 = 4%; *FD* pseudo-R^2 level 2 = 0.02%, level 3 = 2%, level 4 = 3%).

A 10% increase in PUC significantly predicted a 1.8% decrease in *ECVs* values. Compared to community-based samples, population-based samples predicted a significant decrease in *ECVs* (-4%), ω_{HS} (-13%), *H* (-11%, trended to significance), and *FD* values (-6%, trended to significance). Using questionnaires predicted a 7% decrease in *ECVs* values compared to interviews, while combining

multiple measures predicted a 4% increase in *ECVs* values compared to using a single measure. Ordinal indicators predicted a 12% decrease in ω_{HS} values compared to binary indicators.

Regression coefficients were similar after removing outlying and influential cases, but predictors that were on the cusp of our alpha level (.05) changed in significance (see Appendix 1, Tables A1.7 and A1.8).

1.5 Discussion

In this chapter, evaluated the bifactor models of 68 studies using model-based reliability indices to determine how well variation in item responses was represented by general (*p*) and specific psychopathology factors. I found that the *p* factor explained more than half of the modelled variance, but not enough to rule out the importance of specific factors. Furthermore, the *p* factor accounted for most of the variance in raw total scores, while specific factors accounted for around half of the variance in raw subscale scores. The *p* factor and specific factors were reliably specified by their indicators and observed factor scores. In meta-regressions, the percentage of uncontaminated correlations (PUC), measure composition, indicator type, method type, sample type, and estimator significantly predicted variation in the *p* factor's reliability across most indices, as well as the percentage of modelled variance explained by specific factors.

1.5.1 Question 1. Do measurement models of broadband psychopathology measures resemble a bifactor structure?

The modelled variance (*ECV*) in item responses across studies was mainly explained by the *p* factor (61%), but a substantial portion was also explained by the

specific psychology factors (39%). As the modelled variance was distributed between both general and specific sources, it may not be appropriate to exclusively model a single factor that 'unites all disorders' (Caspi & Moffitt, 2018)–this is perhaps more accurate for narrow-band psychology measures (Rodriguez et al., 2016a). Neither is it appropriate to model specific domains without capturing the commonalities among them. Measurement models therefore need to capture the multidimensionality apparent in broadband measures of psychopathology.

Nonetheless, the multidimensionality observed may have arisen from strong specific externalizing factors, and weak specific internalizing factors with items that load preferentially onto the *p* factor (e.g., Calkins et al., 2015; Caspi et al., 2014; Lahey et al., 2012; Laceulle et al., 2015; Liu et al., 2017). The modelled variance may appear to be split between general and specific sources, but it really reflects a two-factor or bifactor *S-1* structure (Eid, Geiser, Koch, & Heene, 2017), with the internalizing dimension disguised as the *p* factor. However, on average, specific internalizing factors explained as much of the modelled variance as externalizing factors, suggesting that multidimensionality was the result of multiple, equally weighted and independent sources.

1.5.2 Question 2. Do total and subscale scores reliably reflect variation in the general and specific psychopathology factors, respectively?

On average, the *p* factor and specific factors explained 94% of the variance in raw total scores (i.e. ω), most of which was explained by the *p* factor (ω_H = 78%). Therefore, individual differences in total scores on broadband psychopathology measures can be considered 'essentially unidimensional', or attributable to a common source (i.e. *p*), despite their multidimensional latent structure. This

discrepancy between measurement in research (multidimensionality) and practice (unidimensionality) may reflect a difficulty in assessing specific domains of a dimensional construct (Gignac, 2014), or the confounding effects of state (or trait) levels of general psychopathology on symptom reporting, similar to the confounding effect of motivation on test taking ability (Duckworth, Quinn, Lynam, Loeber, & Stouthhamer-Loeber, 2011).

The *p* factor and specific psychopathology factors also explained 87% of the variance in raw subscale scores on average (i.e. ω_S). However, only 39% of the variance with error, and 49% without error, was attributable to the specific factors, with the *p* factor explaining the rest. Therefore, the extent to which specific problem domains can be precisely assessed beyond common features in practice is questionable. A similar result was reported by Rodriguez et al. (2016a), where only 27% of the variance in subscale scores on a range of psychological measures could be explained by specific factors beyond the general factor. This is partly because specific factors are specified by fewer items than the general factor and so will naturally contain less information (DeMars, 2013). However, it may be unreasonable to treat subscales as pure indicators of a specific problem domain that are distinct from the overall construct assessed.

1.5.3 Question 3. Can we reliably estimate observed and latent factor scores for the general and specific psychopathology factors?

On average, 92% of variance in the *p* factor could be explained by its indicators, demonstrating high levels of construct reliability (i.e. *H*). Specific factors were also well-represented well by their indicators (*H* = .70). Hence, both the *p* factor and specific factors tend to be reliably specified in structural equation models

using broadband psychopathology measures. This does not mean that the factors estimated across studies are comparable. Rather, the indicators used within each study, on average, represent the general and specific constructs well.

There was also strong overlap between observed *p* factor scores and latent variables (*FD* = .95), and near-acceptable overlap between specific factor scores and latent variables (*FD* = .87). Therefore, one can be relatively confident that they will get similar values each time they estimate general and specific factor scores. Latent variables are preferred for most analyses, but when the option is not available (such as when models are too complex to estimate), clinical researchers can have some confidence in the reliability of factor scores for the *p* factor and specific factors.

1.5.4 Question 4. What methodological characteristics predict variability in the reliability of general and specific psychopathology factors?

Measures with a high PUC (i.e. numerous subscales each with a few items) produced stronger *p* factors and weaker specific factors across all reliability indices. Combining multiple measures weakened the *p* factor reliabilities (ω_H, *H*, and *FD*) and strengthened the specific factor reliabilities (*ECVs*), presumably because it would decrease PUC and introduce method effects. Compared to interview methods, questionnaires favoured *p* factor reliabilities (*ECV*, ω_H) and diminished specific factor reliabilities (*ECVs*), most likely because questionnaires have higher PUC values on average. Overall, a measure's structure is important in affecting the strength of the general and specific psychopathology factors, which are first and foremost methodological entities before they are theoretical constructs.

Ordinal indicators predicted stronger *p* factors and weaker specific factors (*ECV*/*ECVs*, ω/ω_H, *H*). It is uncertain why ordinal indicators, which hold more

information, predicted less reliable specific factors. This may be due to other study characteristics associated with ordinal indicators that favour general factors (e.g., questionnaire methods, higher PUC). Studies that used weighted-least squares estimation showed stronger *p* factors (ω_H, *H*), perhaps because polychoric correlation matrices tend to inflate parameter estimates (Flora & Curran, 2004).

Finally, population-based samples were associated with stronger *p* factors (*ECV*) and weaker specific factors (*ECVs*, ω_H) compared to community and clinical samples. This might be because population samples include a range of respondents, hence responses, which would strengthen the positive manifold among items. Nonetheless, respondents who do not meet the criteria for any disorder in population cohorts might artificially inflate the positive manifold (Watts et al., 2021). Future studies are needed to determine the extent to which substantive and artifactual sample characteristics contribute to the positive manifold underpinning the *p* factor (see Sellbom & Tellegen, 2019, for an example of how a general factor can be underpinned by restricted item responses).

1.5.5 Limitations

A limitation of our study is that there was a substantial degree of between-study heterogeneity which was not entirely explained by methodological differences (at least in the way that I measured this). In some ways this is not surprising; bifactor studies vary widely in their samples, measures, and analyses. However, we can consider whether pooling the reliability estimates was appropriate. Moreover, it may be misguided to refer to 'the' *p* factor as a single construct replicated across studies. Instead, there appear to be multiple *p* factors that differ depending on the sample characteristics, measures, type of

psychopathology assessed, and variations in item loadings (Levin-Aspenson et al., 2020; Watts, Lane, Bonifay, Steinley, & Meyer, 2020; but see Hoffmann et al., 2021).

Another limitation is our judicious use of model-based reliability indices. The formulae for calculating reliability indices assume that the general and specific factors are estimated from a CFA with a simple structure (Rodriguez et al., 2016b). However, some studies included cross-loadings or specific factor correlations. Moreover, model-based reliability indices were designed for continuous outcome variables (Rodriguez et al., 2016b). While they can be applied to categorical outcomes, their reliability will ultimately depend on how skewed the categorical outcome variables are (Flora & Curran, 2004). More than half of studies used weighted least-squares estimation to avoid the distributional problems associated with categorical outcomes. However, calculating reliability estimates from polychoric correlation matrices requires caution as we are no longer estimating the reliability of raw scores, but their hypothesised continuous distributions (Chalmers, 2017).

Finally, I did not control for the quality of model specification. This is complicated by the fact that most, if not all, studies reported acceptable fit for their bifactor models. As I have described, however, model fit statistics are biased indicators of model quality (Greene et al., 2019). Furthermore, most studies estimated *p* factors without considering whether the pattern of loadings resembled a healthy bifactor structure or how they relate to theory. The result is a systematic bias in the studies analysed, reflecting an excitement around bifactor models at the expense of careful analytic practices.

1.5.6 Implications

Research. Bifactor models are a powerful tool for modelling the multidimensionality in psychopathology measures; they offer a solution to the dilemma of lumping or splitting mental health problems by including *both* general *and* specific features within the same model. Nonetheless, clinical scientists need to consider the issues associated with this method, such as its potential to overfit the data and unstable factor loadings (Bornovalova et al., 2020). There are also difficulties in interpreting the *p* factor and specific factors (Sellbom & Tellegen, 2019). For instance, the *p* factor may simply reflect the sum of its parts, rather than a universal trait that is independent of its indicators (Fried, Greene, & Eaton, 2021; Watts et al., 2019).[3] Furthermore, it is uncertain what specific internalizing and externalizing factors distinct from general psychopathology represent (Bonifay, Lane, & Reise, 2017; but see Caspi et al., 2014 for a personality-based interpretation).

Researchers may instead model multidimensionality with the higher-order model, in which the general 'higher-order' psychopathology factor emerges from the correlations among lower-order order psychopathology dimensions such as internalizing and externalizing (Markon, 2019). Some prefer the higher-order model because lower-order factors like internalizing and externalizing appear more stable, interpretable, and consistent over time and methods (Forbes et al., 2021; Funkhouser et al., 2021; Greene et al., 2021; Hyland et al., 2020). The higher-order model is also

[3]We may be able to defend against this issue by sampling more items (Rodriguez et al., 2016a), but until these items are consistently broad across studies, *p* factors will continue to reflect hodgepodges of study-specific items.

consistent with traditional and contemporary dimensional frameworks of psychopathology (Achenbach, 2020; Kotov et al., 2017).

Despite the apparent advantages of higher-order models, I encourage researchers to consider their use carefully. Higher-order models may produce more reliable lower-order dimensions than bifactor models, but their interpretation is equally challenging. Lower-order dimensions are mixtures of variance unique to a problem domain and common to all domains (Gignac, 2008). Therefore, it is unclear how much the lower-order dimensions represent the common variance. In fact, associations between lower-order dimensions and external criteria may be driven by the general factor (Caspi et al., 2014; Lahey et al., 2012). Furthermore, controlling for the general factor uncovers specific treatment effects that would otherwise be masked by the common variance (Aitken et al., 2020; Constantinou et al., 2019; Constantinou et al., 2020; Wade et al., 2018).

There is no clear winner when it comes to modelling the multidimensional structure of psychopathology. However, I would question why there needs to be a winner, when these models are siblings, statistically speaking (van Bork et al., 2017; see also Clark et al., 2021). There are broader issues that affect both models and stifle progress in the field, such as the way modern theories of psychopathology lack clear, falsifiable hypotheses, and are conflated with the statistical models themselves (Fried, 2021). In many ways, we are repeating the challenges of old, whereby the decision between which model to choose ultimately rests on a value judgement (Lilienfeld, 2003), as did the decision of which psychiatric diagnosis to offer or which interpretation of the unconscious to make before that. Until we find consistent ways of measuring the subjective nature of psychopathology indicators, this issue is likely to affect the practice of assessing psychopathology.

Practice. What do our findings suggest for clinicians assessing mental health problems? The high internal consistency of total scores on broadband measures suggests that some general index of mental health can be reliably estimated across various problems domains. Total scores provide more information about mental health functioning than any single disorder or problem domain, making them fitting indicators of prognosis, treatment intensity, and risk (Bach & First, 2018). Total scores may also be used to assess clinical stages of mental health, ranging from no current symptoms to recurrent and severe disorder (Nelson, McGorry, & Fernandez, 2021).

Total scores also compliment transdiagnostic approaches to service delivery, such as the THRIVE framework (Wolpert et al., 2019), Trauma-Informed Care (Sweeney, Clement, Filson, & Kennedy, 2016), Adaptive Mentalization-Based Integrative Treatment (Bevington, Fuggle, Cracknell, & Fonagy, 2017), and Open Dialogue (Seikkula & Olson, 2003). These approaches conceptualize mental health problems in broad terms, such as the degree to which young people, families and communities are thriving in multiple areas of life (e.g., education, vocation, social cohesion), or the way in which trauma has broad effects on social, emotional, and cognitive functioning. Total scores on broadband psychopathology measures or well-being measures might be more suitable than disorder-specific measures for selecting treatment options and assessing therapeutic change under transdiagnostic systems of care. For example, one can imagine that children and adults with low, low-moderate, moderate, and high *p* scores may be suited to Getting Advice (e.g., advice, self-help, sign-posting to relevant services and resources), Getting Help (e.g., brief, goals-focused interventions), Getting More Help (e.g., longer-term interventions), and Getting Risk Support (e.g., risk management, crisis response,

inpatient care), respectively, under the THRIVE framework (Wolpert et al., 2019). We should be mindful that internal consistency (i.e. the degree to which items on a scale are 'summable') is necessary but not sufficient for unidimensionality (i.e. variation in items being underpinned by a single construct or source; Clark & Watson, 2019). Therefore, symptom or disorder ratings may be readily summed to form total scores because of multiple correlated sources rather than a single underlying dimension of mental health.[4] Clinicians who wish to use total scores as a general index of mental health should therefore be aware that this index might differ depending on the problem domains assessed. For example, a young person with particularly high scores on externalizing difficulties might show lower total scores on the Revised Children's Anxiety and Depression Scale (RCADS; Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000) than the Youth Self-Report (YSR; Achenbach & Rescorla, 2001), as the former measure has a weaker representation of externalizing items. Psychometric scores should not be taken as gospel; they are always influenced by the properties of our measure. For instance, the current meta-regressions suggests that a single questionnaire with multiple response options and a high PUC provides a more reliable index of general mental health, than combining multiple questionnaires with binary response options and a low PUC.

Our findings also call into question the reliability of subscale scores in assessing specific problem domains. Clinicians may feel they can measure a specific profile of difficulties using subscales scores, but those scores are partially skewed by global severity or distress. This was illustrated by Thomas (2012), who showed that outpatients with a diagnosis of depression no longer had elevated scores across

[4]This may explain why the *p* factor shifts in its loading patterns depending on which indicators or specific factors are included (Stanton et al., 2021).

problem domains besides depression on the Brief Symptom Inventory (e.g., nervous tension, psychoticism), once general psychopathology was accounted for. Adjusting for the influence of general sources would therefore be important when assessing the specific needs of patients using subscales (e.g., by removing total scores from subscale scores).

It is important to note that specific factors, and hence their respective subscales, are unreliable *relative to* the general factor. A specific factor with 100 healthy loadings might be less reliable than a general factor with 1000 loadings, but that specific factor is likely to show a good degree of reliability in absolute terms. Even the *p* factor shows low reliability when modelled as a specific factor alongside other psychological domains (e.g., personality, personality disorder, and cognitive functioning) and a "Big Everything" general factor (Littlefield, Lane, Gette, Watts, & Sher, 2021). Therefore, low omega reliability does not preclude the use of subscales, but care should be taken to ensure that the specific domain is sampled by a variety of well-worded and diverse items (Clark & Watson, 2019; Watts, Boness, Loeffelman, Steinley, & Sher, 2021).

1.5.7 Conclusion

Bifactor studies of broadband assessment measures generally support the multidimensional nature of psychopathology. Both general and specific features of common mental health problems can be reliably assessed using self-report measures. However, the extent to which we measure general over specific features depends on the way scores are treated (e.g., optimally-weighted or item-weighted) and the methods used (e.g., questionnaire vs interview). This calls for a pragmatic approach to mental health assessment, where different measures or methods are

required for different purposes (e.g., research studies, clinical assessment, different types of problem domains, Achenbach, 2021; Markon, 2021) rather than a single, all-purpose measure (Lahey, Moore, Kaczkurkin, & Zald, 2021).

1.6 References

Achenbach, T. M. (2020). Bottom-Up and Top-Down Paradigms for Psychopathology: A Half-Century Odyssey. *Annu Rev Clin Psychol, 16*, 1-24. doi:10.1146/annurev-clinpsy-071119-115831

Achenbach, T. M. (2021). Hierarchical dimensional models of psychopathology: yes, but. *World Psychiatry, 20*(1), 64-65. doi:10.1002/wps.20810

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

Afzali, M. H., Sunderland, M., Carragher, N., & Conrod, P. (2018). The Structure of Psychopathology in Early Adolescence: Study of a Canadian Sample: La structure de la psychopathologie au debut de l'adolescence: etude d'un echantillon canadien. *Can J Psychiatry, 63*(4), 223-230. doi:10.1177/0706743717737032

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. *J Child Psychol Psychiatry, 61*(9), 998-1008. doi:10.1111/jcpp.13194

Arrindell, W. A., Urban, R., Carrozzino, D., Bech, P., Demetrovics, Z., & Roozen, H. G. (2017). SCL-90-R emotional distress ratings in substance use and impulse control disorders: One-factor, oblique first-order, higher-order, and bi-factor

models compared. *Psychiatry Res, 255*, 173-185. doi:10.1016/j.psychres.2017.05.019

Bach, B., & First, M. B. (2018). Application of the ICD-11 classification of personality disorders. *BMC Psychiatry, 18*(1), 351. doi:10.1186/s12888-018-1908-3

Black, L., Panayiotou, M., & Humphrey, N. (2019). The dimensionality and latent structure of mental health difficulties and wellbeing in early adolescence. *PLoS One, 14*(2), e0213018. doi:10.1371/journal.pone.0213018

Blazquez-Rincon, D., Duran, J. I., & Botella, J. (2021). The Fear of COVID-19 Scale: A Reliability Generalization Meta-Analysis. *Assessment*, 1073191121994164. doi:10.1177/1073191121994164

Bonett, D. G. (2002). Sample Size Requirements for Testing and Estimating Coefficient Alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335-340. doi:10.3102/10769986027004335

Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychol Methods, 15*(4), 368-385. doi:10.1037/a0020142

Bonifay, W., Lane, S. P., & Reise, S. P. (2016). Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology. *Clinical Psychological Science, 5*(1), 184-186. doi:10.1177/2167702616657069

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods, 1*(2), 97-111. doi:10.1002/jrsm.12

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. *Biol Psychiatry, 88*(1), 18-27. doi:10.1016/j.biopsych.2020.01.013

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One, 6*(11), e27407. doi:10.1371/journal.pone.0027407

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. doi:10.1177/2167702619859332

Brodbeck, J., Stulz, N., Itten, S., Regli, D., Znoj, H., & Caspar, F. (2014). The structure of psychopathological symptoms and the associations with DSM-diagnoses in treatment seeking individuals. *Compr Psychiatry, 55*(3), 714-726. doi:10.1016/j.comppsych.2013.11.001

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., . . . Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry, 56*(12), 1356-1369. doi:10.1111/jcpp.12416

Carragher, N., Teesson, M., Sunderland, M., Newton, N. C., Krueger, R. F., Conrod, P. J., . . . Slade, T. (2016). The structure of adolescent psychopathology: a symptom-level analysis. *Psychol Med, 46*(5), 981-994. doi:10.1017/S0033291715002470

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci, 2*(2), 119-137. doi:10.1177/2167702613497473

Caspi, A., & Moffitt, T. E. (2018). All for One and One for All: Mental Disorders in One Dimension. *Am J Psychiatry, 175*(9), 831-844. doi:10.1176/appi.ajp.2018.17121383

Castellanos-Ryan, N., Briere, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., . . . Consortium, I. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *J Abnorm Psychol, 125*(8), 1039-1052. doi:10.1037/abn0000193

Castellanos-Ryan, N., Briere, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., . . . Consortium, I. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *J Abnorm Psychol, 125*(8), 1039-1052. doi:10.1037/abn0000193

Cervin, M., Norris, L. A., Ginsburg, G., Gosch, E. A., Compton, S. N., Piacentini, J., . . . Kendall, P. C. (2021). The p Factor Consistently Predicts Long-Term Psychiatric and Functional Outcomes in Anxiety-Disordered Youth. *J Am Acad Child Adolesc Psychiatry, 60*(7), 902-912 e905. doi:10.1016/j.jaac.2020.08.440

Chalmers, R. P. (2018). On Misconceptions and the Limited Usefulness of Ordinal Alpha. *Educ Psychol Meas, 78*(6), 1056-1071. doi:10.1177/0013164417727036

Chorpita, B. F., Yim, L. M., Moffitt, C. E., Umemoto L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy*, 38, 835-855.

Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., . . . Sripada, C. (2021). The General Factor of Psychopathology in the Adolescent Brain Cognitive Development (ABCD) Study: A Comparison of Alternative

Modeling Approaches. *Clin Psychol Sci, 9*(2), 169-182. doi:10.1177/2167702620959317

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychol Assess, 31*(12), 1412-1427. doi:10.1037/pas0000626

Constantinou, M. P., & Fonagy, P. (2019). Evaluating bifactor models of psychopathology using model-based reliability indices. *PsyArXiv Preprints*. doi: 10.31234/osf.io/6tf7j

Constantinou, M. P., Frueh, B. C., Fowler, J. C., Allen, J. G., Madan, A., Oldham, J. M., & Fonagy, P. (2020). Predicting depression outcomes throughout inpatient treatment using the general and specific personality disorder factors. *Psychol Med*, 1-9. doi:10.1017/S003329172000361X

Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., . . . Fonagy, P. (2019). Changes in General and Specific Psychopathology Factors Over a Psychosocial Intervention. *J Am Acad Child Adolesc Psychiatry, 58*(8), 776-786. doi: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. *Psychol Assess, 31*(6), 730-740. doi:10.1037/pas0000695

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *J Consult Clin Psychol, 88*(3), 179-195. doi:10.1037/ccp0000482

DeMars, C. E. (2013). A Tutorial on Interpreting Bifactor Model Scores. *International Journal of Testing, 13*(4), 354-378. doi:10.1080/15305058.2013.799067

Deutz, M. H. F., Shi, Q., Vossen, H. G. M., Huijding, J., Prinzie, P., Dekovic, M., . . . Woltering, S. (2018). Evaluation of the Strengths and Difficulties

Questionnaire-Dysregulation Profile (SDQ-DP). *Psychol Assess, 30*(9), 1174-1185. doi:10.1037/pas0000564

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., . . . Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27-46. doi:10.1111/j.1600-0587.2012.07348.x

Du Rietz, E., Pettersson, E., Brikell, I., Ghirardi, L., Chen, Q., Hartman, C., . . . Kuja-Halkola, R. (2021). Overlap between attention-deficit hyperactivity disorder and neurodevelopmental, externalising and internalising disorders: separating unique from general psychopathology effects. *Br J Psychiatry, 218*(1), 35-42. doi:10.1192/bjp.2020.152

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proc Natl Acad Sci U S A, 108*(19), 7716-7720. doi:10.1073/pnas.1018601108

Eaton, N. R., Krueger, R. F., Markon, K. E., Keyes, K. M., Skodol, A. E., Wall, M., . . . Grant, B. F. (2013). The structure and predictive validity of the internalizing disorders. *J Abnorm Psychol, 122*(1), 86-92. doi:10.1037/a0029598

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychol Methods, 22*(3), 541-562. doi:10.1037/met0000083

Etkin, P., Ibáñez, M. I., Ortet, G., & Mezquita, L. (2021). Longitudinal Associations Between the Five-Factor Model of Personality and The Bi-Factor Model of Psychopathology: Continuity, Pathoplasty and Complication Effects in Adolescents. *Journal of Psychopathology and Behavioral Assessment*. doi:10.1007/s10862-021-09903-1

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods, 9*(4), 466-491. doi:10.1037/1082-989X.9.4.466

Forbes, M. K., Greene, A. L., Levin-Aspenson, H. F., Watts, A. L., Hallquist, M., Lahey, B. B., . . . Krueger, R. F. (2021). Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. *J Abnorm Psychol, 130*(3), 297-317. doi:10.1037/abn0000533

Fried, E. I. (2021). Theories and Models: What They Are, What They Are for, and What They Are About. *Psychological Inquiry, 31*(4), 336-344. doi:10.1080/1047840x.2020.1854011

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), 69-70. doi:10.1002/wps.20814

Funkhouser, C. J., Correa, K. A., Letkiewicz, A. M., Cozza, E. M., Estabrook, R., & Shankman, S. A. (2021). Evaluating the criterion validity of hierarchical psychopathology dimensions across models: Familial aggregation and associations with research domain criteria (sub)constructs. *J Abnorm Psychol, 130*(6), 575-586. doi:10.1037/abn0000687

Geeraerts, S. B., Deutz, M. H., Dekovic, M., Bunte, T., Schoemaker, K., Espy, K. A., . . . Matthys, W. (2015). The Child Behavior Checklist Dysregulation Profile in Preschool Children: A Broad Dysregulation Syndrome. *J Am Acad Child Adolesc Psychiatry, 54*(7), 595-602 e592. doi:10.1016/j.jaac.2015.04.012

Gibbons, R. D., Rush, A. J., & Immekus, J. C. (2009). On the psychometric validity of the domains of the PDSQ: an illustration of the bi-factor item response theory model. *J Psychiatr Res, 43*(4), 401-410. doi:10.1016/j.jpsychires.2008.04.013

Gignac, G. E. (2014). Dynamic mutualism versus g factor theory: An empirical test. *Intelligence, 42*, 89-97. doi:10.1016/j.intell.2013.11.004

Gluschkoff, K., Jokela, M., & Rosenstrom, T. (2019). The General Psychopathology Factor: Structural Stability and Generalizability to Within-Individual Changes. *Front Psychiatry, 10*, 594. doi:10.3389/fpsyt.2019.00594

Gomez, R., Stavropoulos, V., Zarate, D., & Palikara, O. (2021). Symptom Checklist-90-Revised: A structural examination in relation to family functioning. *PLoS One, 16*(3), e0247902. doi:10.1371/journal.pone.0247902

Greco, L. M., O'Boyle, E. H., Cockburn, B. S., & Yuan, Z. (2018). Meta-Analysis of Coefficient Alpha: A Reliability Generalization Study. *Journal of Management Studies, 55*(4), 583-618. doi:10.1111/joms.12328

Greene, A. L., & Eaton, N. R. (2017). The temporal stability of the bifactor model of comorbidity: An examination of moderated continuity pathways. *Compr Psychiatry, 72*, 74-82. doi:10.1016/j.comppsych.2016.09.010

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., . . . Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *J Abnorm Psychol, 128*(7), 740-764. doi:10.1037/abn0000434

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430-450. doi:10.1037/1082-989x.6.4.430

Haltigan, J. D., Aitken, M., Skilling, T., Henderson, J., Hawke, L., Battaglia, M., . . . Andrade, B. F. (2018). "P" and "DP:" Examining Symptom-Level Bifactor Models of Psychopathology and Dysregulation in Clinically Referred Children and Adolescents. *J Am Acad Child Adolesc Psychiatry, 57*(6), 384-396. doi:10.1016/j.jaac.2018.03.010

Hankin, B. L., Davis, E. P., Snyder, H., Young, J. F., Glynn, L. M., & Sandman, C. A. (2017). Temperament factors and dimensional, latent bifactor models of child psychopathology: Transdiagnostic and specific associations in two youth samples. *Psychiatry Res, 252*, 139-146. doi:10.1016/j.psychres.2017.02.061

Harden, K. P., Engelhardt, L. E., Mann, F. D., Patterson, M. W., Grotzinger, A. D., Savicki, S. L., . . . Tucker-Drob, E. M. (2020). Genetic Associations Between Executive Functions and a General Factor of Psychopathology. *J Am Acad Child Adolesc Psychiatry, 59*(6), 749-758. doi:10.1016/j.jaac.2019.05.006

Hartung, J. (1999). An Alternative Method for Meta-Analysis. *Biometrical Journal, 41*(8), 901-916. doi:10.1002/(sici)1521-4036(199912)41:8<901::aid-bimj901>3.0.co;2-w

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557-560. doi:10.1136/bmj.327.7414.557

Holzinger, K. J., & Swineford, F. (1937). The Bi-factor method. *Psychometrika, 2*(1), 41-54. doi:10.1007/bf02287965

Hyland, P., Murphy, J., Shevlin, M., Bentall, R. P., Karatzias, T., Ho, G. W. K., . . . McElroy, E. (2021). On top or underneath: where does the general factor of psychopathology fit within a dimensional model of psychopathology? *Psychol Med, 51*(14), 2422-2432. doi:10.1017/S003329172000104X

Hyland, P., Murphy, J., Shevlin, M., Carey, S., Vallieres, F., Murphy, D., & Elklit, A. (2018). Correlates of a general psychopathology factor in a clinical sample of childhood sexual abuse survivors. *J Affect Disord, 232*, 109-115. doi:10.1016/j.jad.2018.02.048

Ignatyev, Y., Baggio, S., & Mundt, A. P. (2019). The Underlying Structure of Comorbid Mental Health and Substance Use Disorders in Prison Populations. *Psychopathology, 52*(1), 2-9. doi:10.1159/000495844

King, S. M., Saunders, G. R. B., Elkins, I. J., Winters, K. C., Iacono, W. G., & McGue, M. (2020). Where Do Gambling Problems Fit in the Structure of Psychopathology During Emerging Adulthood? *Int Gambl Stud, 20*(1), 1-13. doi:10.1080/14459795.2019.1643901

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol, 126*(4), 454-477. doi:10.1037/abn0000258

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. *J Abnorm Psychol, 116*(4), 645-666. doi:10.1037/0021-843X.116.4.645

Laceulle, O. M., Vollebergh, W. A. M., & Ormel, J. (2015). The Structure of Psychopathology in Adolescence. *Clinical Psychological Science, 3*(6), 850-860. doi: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? *J Abnorm Psychol, 121*(4), 971-977. doi:10.1037/a0028355

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychol Bull, 143*(2), 142-186. doi:10.1037/bul0000069

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: empirical support, implications, and remaining issues. *World Psychiatry, 20*(1), 57-63. doi: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. *J Child Psychol Psychiatry, 56*(4), 415-422. doi:10.1111/jcpp.12300

Lahey, B. B., Zald, D. H., Perkins, S. F., Villalta-Gil, V., Werts, K. B., Van Hulle, C. A., . . . Waldman, I. D. (2018). Measuring the hierarchical general factor model of psychopathology in young adults. *Int J Methods Psychiatr Res, 27*(1). doi:10.1002/mpr.1593

Langan, D., Higgins, J. P. T., & Simmonds, M. (2017). Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies. *Res Synth Methods, 8*(2), 181-198. doi:10.1002/jrsm.1198

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. doi:10.1177/1073191120954921

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: reflections and directions. *J Abnorm Child Psychol, 31*(3), 285-291. doi:10.1023/a:1023229529866

Littlefield, A. K., Lane, S. P., Gette, J. A., Watts, A. L., & Sher, K. J. (2021). The "Big Everything": Integrating and investigating dimensional models of psychopathology, personality, personality pathology, and cognitive functioning. *Personal Disord, 12*(2), 103-114. doi:10.1037/per0000457

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-

categorical measures. *Psychol Methods, 22*(3), 486-506. doi:10.1037/met0000075

Livesley, W. J. (2011). An empirically-based classification of personality disorder. *Journal of Personality Disorders*, *25*(3), 397.

Lopez-Pina, J. A., Sanchez-Meca, J., Lopez-Lopez, J. A., Marin-Martinez, F., Nunez-Nunez, R. M., Rosa-Alcazar, A. I., . . . Ferrer-Requena, J. (2015). The Yale-Brown Obsessive Compulsive Scale: A Reliability Generalization Meta-Analysis. *Assessment, 22*(5), 619-628. doi:10.1177/1073191114551954

Lowe, B., Spitzer, R. L., Williams, J. B., Mussell, M., Schellberg, D., & Kroenke, K. (2008). Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry, 30*(3), 191-199. doi:10.1016/j.genhosppsych.2008.01.001

Mann, F. D., Atherton, O. E., DeYoung, C. G., Krueger, R. F., & Robins, R. W. (2020). Big five personality traits and common mental disorders within a hierarchical taxonomy of psychopathology: A longitudinal study of Mexican-origin youth. *J Abnorm Psychol, 129*(8), 769-787. doi:10.1037/abn0000633

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol, 15*, 51-69. doi:10.1146/annurev-clinpsy-050718-095522

Markon, K. E. (2021). On hierarchically-informed measures of psychopathology. *World Psychiatry, 20*(1), 66-67. doi:10.1002/wps.20812

Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32*(1), 151-170. doi:10.1016/j.cedpsych.2006.10.008

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosario, M. C., Mari, J. J., . . . Salum, G. A. (2017). A general psychopathology factor (P factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *J Abnorm Psychol, 126*(1), 137-148. doi:10.1037/abn0000205

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? *J Child Psychol Psychiatry, 59*(6), 667-675. doi:10.1111/jcpp.12849

Miller, M., Iosif, A. M., Young, G. S., Bell, L. J., Schwichtenberg, A. J., Hutman, T., & Ozonoff, S. (2019). The dysregulation profile in preschoolers with and without a family history of autism spectrum disorder. *J Child Psychol Psychiatry, 60*(5), 516-523. doi:10.1111/jcpp.13003

Moberget, T., Alnaes, D., Kaufmann, T., Doan, N. T., Cordova-Palomera, A., Norbom, L. B., . . . Westlye, L. T. (2019). Cerebellar Gray Matter Volume Is Associated With Cognitive Function and Psychopathology in Adolescence. *Biol Psychiatry, 86*(1), 65-75. doi:10.1016/j.biopsych.2019.01.019

Moore, T. M., Kaczkurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., . . . Lahey, B. B. (2020). doi:10.1101/2020.04.27.064303

Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The Development of the General Factor of Psychopathology 'p Factor' Through Childhood and Adolescence. *J Abnorm Child Psychol, 44*(8), 1573-1586. doi:10.1007/s10802-016-0132-1

Nelson, B., McGorry, P. D., & Fernandez, A. V. (2021). Integrating clinical staging and phenomenological psychopathology to add depth, nuance, and utility to

clinical phenotyping: a heuristic challenge. *The Lancet Psychiatry, 8*(2), 162-168. doi:10.1016/s2215-0366(20)30316-3

Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., . . . Tiemeier, H. (2016). Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. *J Am Acad Child Adolesc Psychiatry, 55*(12), 1038-1045 e1034. doi:10.1016/j.jaac.2016.09.498

Niarchou, M., Moore, T. M., Tang, S. X., Calkins, M. E., McDonald-McGuinn, D. M., Zackai, E. H., . . . Gur, R. E. (2017). The dimensional structure of psychopathology in 22q11.2 Deletion Syndrome. *J Psychiatr Res, 92*, 124-131. doi:10.1016/j.jpsychires.2017.04.006

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A Heuristic for Developing Transdiagnostic Models of Psychopathology: Explaining Multifinality and Divergent Trajectories. *Perspect Psychol Sci, 6*(6), 589-609. doi:10.1177/1745691611419672

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. *J Abnorm Child Psychol, 46*(7), 1373-1383. doi:10.1007/s10802-018-0398-6

Oro, V., Goldsmith, H. H., & Lemery-Chalfant, K. (2021). Elucidating the Links Between Mother and Father Alcohol Use Disorder and Adolescent Externalizing Psychopathology: A Test of Transmission Specificity Within Competing Factor Structures and Genetic and Environmental Liabilities. *Behav Genet, 51*(5), 512-527. doi:10.1007/s10519-021-10072-w

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *Br J Psychiatry, 207*(1), 15-22. doi:10.1192/bjp.bp.114.149591

Patterson, V. C., Pencer, A., Pavlova, B., Awadia, A., MacKenzie, L. E., Zwicker, A., . . . Uher, R. (2021). Youth Experience Tracker Instrument: A self-report measure of developmental antecedents to severe mental illness. *Early Interv Psychiatry, 15*(3), 676-685. doi:10.1111/eip.13007

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. *J Am Acad Child Adolesc Psychiatry, 57*(6), 372-383. doi:10.1016/j.jaac.2017.12.016

Pezzoli, P., Antfolk, J., & Santtila, P. (2017). Phenotypic factor analysis of psychopathology reveals a new body-related transdiagnostic factor. *PLoS One, 12*(5), e0177674. doi:10.1371/journal.pone.0177674

Preti, A., Carta, M. G., & Petretto, D. R. (2019). Factor structure models of the SCL-90-R: Replicability across community samples of adolescents. *Psychiatry Res, 272*, 491-498. doi:10.1016/j.psychres.2018.12.146

Reise, S. P. (2012). Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behav Res, 47*(5), 667-696. doi: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. *J Pers Assess, 95*(2), 129-140. doi:10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Assess, 92*(6), 544-559. doi:10.1080/00223891.2010.496477

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2012). Multidimensionality and Structural Coefficient Bias in Structural Equation

Modeling. *Educational and Psychological Measurement, 73*(1), 5-26. doi:10.1177/0013164412449831

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *J Pers Assess, 98*(3), 223-237. doi:10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol Methods, 21*(2), 137-150. doi:10.1037/met0000045

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychol Methods, 11*(3), 306-322. doi:10.1037/1082-989X.11.3.306

Romani-Sponchiado, A., Vidal-Ribas, P., Bressan, R. A., de Jesus Mari, J., Miguel, E. C., Gadelha, A., . . . Hoffmann, M. S. (2021). Longitudinal associations between positive attributes and psychopathology and their interactive effects on educational outcomes. *Eur Child Adolesc Psychiatry*. doi:10.1007/s00787-021-01882-8

Romer, A. L., Knodt, A. R., Houts, R., Brigidi, B. D., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2018). Structural alterations within cerebellar circuitry are associated with general liability for common mental disorders. *Mol Psychiatry, 23*(4), 1084-1090. doi:10.1038/mp.2017.57

Romer, A. L., Knodt, A. R., Sison, M. L., Ireland, D., Houts, R., Ramrakha, S., . . . Hariri, A. R. (2021). Replicability of structural brain alterations associated with general psychopathology: evidence from a population-representative birth cohort. *Mol Psychiatry, 26*(8), 3839-3846. doi:10.1038/s41380-019-0621-z

Rosenstrom, T., Gjerde, L. C., Krueger, R. F., Aggen, S. H., Czajkowski, N. O., Gillespie, N. A., . . . Ystrom, E. (2019). Joint factorial structure of

psychopathology and personality. *Psychol Med, 49*(13), 2158-2167. doi:10.1017/S0033291718002982

Rubio-Aparicio, M., Nunez-Nunez, R. M., Sanchez-Meca, J., Lopez-Pina, J. A., Marin-Martinez, F., & Lopez-Lopez, J. A. (2020). The Padua Inventory-Washington State University Revision of Obsessions and Compulsions: A Reliability Generalization Meta-Analysis. *J Pers Assess, 102*(1), 113-123. doi:10.1080/00223891.2018.1483378

Rytila-Manninen, M., Frojd, S., Haravuori, H., Lindberg, N., Marttunen, M., Kettunen, K., & Therman, S. (2016). Psychometric properties of the Symptom Checklist-90 in adolescent psychiatric inpatients and age- and gender-matched community youth. *Child Adolesc Psychiatry Ment Health, 10*, 23. doi:10.1186/s13034-016-0111-x

Sallis, H., Szekely, E., Neumann, A., Jolicoeur-Martineau, A., van, I. M., Hillegers, M., . . . Evans, J. (2019). General psychopathology, internalising and externalising in children and functional outcomes in late adolescence. *J Child Psychol Psychiatry, 60*(11), 1183-1190. doi:10.1111/jcpp.13067

Sanchez-Meca, J., Lopez-Lopez, J. A., & Lopez-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *Br J Math Stat Psychol, 66*(3), 402-425. doi:10.1111/j.2044-8317.2012.02057.x

Schaefer, J. D., Moffitt, T. E., Arseneault, L., Danese, A., Fisher, H. L., Houts, R., . . . Caspi, A. (2018). Adolescent Victimization and Early-Adult Psychopathology: Approaching Causal Inference Using a Longitudinal Twin Study to Rule Out Noncausal Explanations. *Clin Psychol Sci, 6*(3), 352-371. doi:10.1177/2167702617741381

Seikkula, J., & Olson, M. E. (2003). The open dialogue approach to acute psychosis: its poetics and micropolitics. *Fam Process, 42*(3), 403-418. doi:10.1111/j.1545-5300.2003.00403.x

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychol Assess, 31*(12), 1428-1441. doi:10.1037/pas0000623

Sheldrick, R. C., Henson, B. S., Merchant, S., Neger, E. N., Murphy, J. M., & Perrin, E. C. (2012). The Preschool Pediatric Symptom Checklist (PPSC): development and initial validation of a new social/emotional screening instrument. *Acad Pediatr, 12*(5), 456-467. doi:10.1016/j.acap.2012.06.008

Shields, A. N., Reardon, K. W., Brandes, C. M., & Tackett, J. L. (2019). The p factor in children: Relationships with executive functions and effortful control. *Journal of Research in Personality, 82*, 103853. doi:10.1016/j.jrp.2019.103853

Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The General Factor of Psychopathology. *Annu Rev Clin Psychol, 16*, 75-98. doi: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. *Clin Psychol Sci, 5*(1), 98-110. doi:10.1177/2167702616651076

Spearman, C. (1904). "General intelligence" Objectively Determined and Measured. *American Journal of Psychology, 15*(2), 201–293. doi:10.2307/1412107

St Clair, M. C., Neufeld, S., Jones, P. B., Fonagy, P., Bullmore, E. T., Dolan, R. J., . . . Goodyer, I. M. (2017). Characterising the latent structure and organisation of self-reported thoughts, feelings and behaviours in adolescents and young adults. *PLoS One, 12*(4), e0175381. doi:10.1371/journal.pone.0175381

Stochl, J., Khandaker, G. M., Lewis, G., Perez, J., Goodyer, I. M., Zammit, S., . . . Jones, P. B. (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychol Med, 45*(7), 1483-1493. doi:10.1017/S003329171400261X

Sweeney, A., Clement, S., Filson, B., & Kennedy, A. (2016). Trauma-informed mental healthcare in the UK: what is it and how can we further its development? *Mental Health Review Journal, 21*(3), 174-192. doi:10.1108/mhrj-01-2015-0006

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. *J Abnorm Psychol, 122*(4), 1142-1153. doi:10.1037/a0034151

Tang, A., Crawford, H., Morales, S., Degnan, K. A., Pine, D. S., & Fox, N. A. (2020). Infant behavioral inhibition predicts personality and social outcomes three decades later. *Proc Natl Acad Sci U S A, 117*(18), 9800-9807. doi:10.1073/pnas.1917376117

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69*(4), 613-625. doi:10.1007/bf02289858

Thomas, M. L. (2012). Rewards of bridging the divide between measurement and clinical theory: demonstration of a bifactor model for the Brief Symptom Inventory. *Psychol Assess, 24*(1), 101-113. doi:10.1037/a0024712

Thompson, E. J., Richards, M., Ploubidis, G. B., Fonagy, P., & Patalay, P. (2021). Changes in the adult consequences of adolescent mental ill-health: findings from the 1958 and 1970 British birth cohorts. *Psychol Med*, 1-10. doi:10.1017/S0033291721002506

Thurstone, L. L. (1944). Second-order factors. *Psychometrika*, *9*(2), 71-100.

Urban, R., Arrindell, W. A., Demetrovics, Z., Unoka, Z., & Timman, R. (2016). Cross-cultural confirmation of bi-factor models of a symptom distress measure: Symptom Checklist-90-Revised in clinical samples. *Psychiatry Res, 239*, 265-274. doi:10.1016/j.psychres.2016.03.039

Urban, R., Kun, B., Farkas, J., Paksi, B., Kokonyei, G., Unoka, Z., . . . Demetrovics, Z. (2014). Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. *Psychiatry Res, 216*(1), 146-154. doi:10.1016/j.psychres.2014.01.027

Vacha-Haase, T. (2016). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. *Educational and Psychological Measurement, 58*(1), 6-20. doi:10.1177/0013164498058001002

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology, 27*(6), 759-773. doi:10.1177/0959354317737185

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., . . . Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods, 7*(1), 55-79. doi:10.1002/jrsm.1164

Vosberg, D. E., Syme, C., Parker, N., Richer, L., Pausova, Z., & Paus, T. (2021). Sex continuum in the brain and body during adolescence and psychological traits. *Nat Hum Behav, 5*(2), 265-272. doi:10.1038/s41562-020-00968-8

Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2018). Effect of Foster Care Intervention on Trajectories of General and Specific Psychopathology Among Children With Histories of Institutional Rearing: A Randomized Clinical Trial. *JAMA Psychiatry, 75*(11), 1137-1145. doi:10.1001/jamapsychiatry.2018.2556

Wade, M., Plamondon, A., & Jenkins, J. M. (2021). A Family Socialization Model of Transdiagnostic Risk for Psychopathology in Preschool Children. *Res Child Adolesc Psychopathol, 49*(8), 975-988. doi:10.1007/s10802-021-00789-x

Watts, A. L., Boness, C. L., Loeffelman, J. E., Steinley, D., & Sher, K. J. (2021). Does crude measurement contribute to observed unidimensionality of psychological constructs? A demonstration with DSM-5 alcohol use disorder. *J Abnorm Psychol, 130*(5), 512-524. doi:10.1037/abn0000678

Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. C. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychol Inq, 31*(4), 310-320. doi: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. doi:10.1177/2167702619855035

Weissman, D. G., Bitran, D., Miller, A. B., Schaefer, J. D., Sheridan, M. A., & McLaughlin, K. A. (2019). Difficulties with emotion regulation as a transdiagnostic mechanism linking child maltreatment with the emergence of psychopathology. *Dev Psychopathol, 31*(3), 899-915. doi:10.1017/S0954579419000348

Weissman, D. G., Nook, E. C., Dews, A. A., Miller, A. B., Lambert, H. K., Sasse, S. F., . . . McLaughlin, K. A. (2020). Low Emotional Awareness as a Transdiagnostic Mechanism Underlying Psychopathology in Adolescence. *Clin Psychol Sci, 8*(6), 971-988. doi:10.1177/2167702620923649

Part 2. Empirical Paper

The Roles of Stressful Life Events and Family Obligation in the Relationship Between Socioeconomic Status and the General and Specific Psychopathology Factors

2.1 Abstract

In this chapter, I aim to explore developmental questions related to the bifactor model of psychopathology. Specifically, I investigate the link between socioeconomic status (SES) and the general and specific psychopathology factors in childhood. Socioeconomic disadvantage appears to predict child mental health outcomes partly through an increased incidence of stressful life events (SLEs), e.g., housing problems or family conflict. However, it is uncertain whether SLEs pose broad or specific risks to child mental health problems. Moreover, children may differ in how sensitive they are to the impact of stressful life events, based on how dependent they are on their families' views and values (i.e. family obligation). I estimated a bifactor model, with a general *p* factor and specific internalizing, externalizing, and attention factors, using child-reported emotional, behavioural, and attentional symptoms in 10,173 community-based 9-14 year-olds recruited in the Adolescent Brain and Cognitive Development (ABCD) study. I used a second-order moderated mediation model to evaluate the role of SLEs in mediating the relationship between socioeconomic status and the general and specific psychopathology factors, and the role of family obligation in moderating the indirect effect of stressful life events on psychopathology factors. Socioeconomic status negatively predicted the *p* factor and specific externalizing factor. SLEs significantly mediated the link between socioeconomic status and the *p* factor, specific externalizing factor, and specific attention factor, explaining 36%, 14%, and 41% of their relationship with SES, respectively. The indirect effect of SES on the *p* factor and specific externalizing and attention factors via SLEs was stronger at higher levels of family obligation, even though higher family obligation levels predicted lower psychopathology scores on average. These findings demonstrate

that SES has both broad and specific influences on child mental health outcomes via SLEs, and that family obligation protects against and bolsters this relationship.

2.2 Introduction

Few would disagree that the environment plays an indisputable role in child development (Bronfenbrenner & Crouter, 1983). Socioeconomic status (SES) reflects differences in children and families' access to environmental resources that promote health, including material capital (e.g., money), social capital (e.g., education, social support), welfare, and clean and safe environments (Shavers, 2007). Despite its ubiquity, most studies control for the effects of SES rather than examine its unique contributions to mental health outcomes. Studies that have examined the role of SES report consistent but modest negative associations with childhood psychopathology, suggesting SES has a distal influence on childhood outcomes (Bradley & Corwyn, 2002; Peverill et al., 2021; Reiss, 2013). Understanding the mechanisms that link SES and child development would not only advance the field by linking macro- and micro-level processes but may also help determine targets for prevention at the population level.

One way in which socioeconomic disadvantage predicts an increase in childhood mental health problems is through an increased risk of stressful life events (SLEs), such as family conflict, peer victimization, housing problems, and neighourhood violence (Reiss et al., 2019). SLEs partially or fully mediate the link between SES and child and adolescent mental health outcomes (Amone-P'olak et al., 2009; Felner et al., 1995; Kang et al., 2011; Ziebold et al., 2021). Family-related SLEs (e.g., family conflict, parent mental health problems, parenting difficulties) have been studied most and consistently mediate the link between SES and mental health outcomes, though their role may decrease with age as adolescents separate from the family unit (Devenish et al., 2017; González et al., 2021; Tracy, Zimmerman, Galea, McCauley, & Stoep, 2008).

SLEs appear to act as proximal risk factors that link the distal influence of SES with child mental health outcomes. However, it is unclear whether SLEs have their greatest impact on internalizing problems (Amone-P'olak et al., 2009), externalizing problems (Ziebold et al., 2021), or general functioning (González et al., 2021; Felner et al., 1995). In psychopathology research, the emphasis is shifting from measuring disorder-specific problems to hierarchically organized transdiagnostic factors (Kotov et al., 2021). There has been a recent focus on factors at the top of the hierarchy, i.e. the general psychopathology (*p*) factor, which accounts for the fact that internalizing and externalizing problems tend to co-occur and may share a common aetiology (Caspi et al., 2014; Lahey et al., 2012). The *p* factor summarises the severity of, and perhaps vulnerability towards, general dysregulation in the form of emotional, behavioural and social difficulties in children and adolescents (Brandes, Herzhoff, Smack, & Tackett, 2019**;** Deutz et al., 2020; Haltigan et al., 2018).

Using a bifactor model, the *p* factor can be estimated alongside specific psychopathology factors like internalizing and externalizing (Markon, 2019). Specific psychopathology factors are orthogonal to the general *p* factor, meaning one can examine the unique associations of internalizing and externalizing, free from the influence of the general variance explained by the *p* factor (Bornavalova, Choate, Fatimah, Petersen, Wiernik, 2020). This would be useful for determining the extent to which SES predicts a broad dysregulation profile compared to specific difficulties. The *p* factor is negatively associated with markers of SES over and above specific internalizing and externalizing factors (Belsky et al., 2019; Blanco et al., 2021; Patalay et al., 2015; Schäfer et al., 2018; Snyder, Young, & Hankin, 2017; Wade, Fox, Zeanah, & Nelson, 2018). In fact, some markers of SES are no longer associated with specific internalizing and externalizing factors once their associations with the *p*

factor are accounted for (Caspi et al., 2014; Lahey et al., 2012), suggesting that socioeconomic disadvantage confers a broad risk to psychopathology.

Two studies investigating SES and the bifactor model are of particular interest. In a sample of disadvantage youth from African-American backgrounds, Liu, Mustanski, Dick, Bolland, and Kertes (2017) found that SLEs and racial discrimination positively predicted the *p* factor, whilst SLEs and exposure to violence positively predicted the specific externalizing factor (associations with specific internalizing problems were not significant). Wade et al. (2021) found that the negative associations between SES markers (family income/assets and maternal education) at two months old and the *p* factor estimated at 36 months was partially mediated by maternal responsiveness at 18 months. That is, mothers' responsiveness to their children whilst playing and reading was one way in which material and social capital was associated with a reduced risk of common emotional and behavioural problems in children. Maternal education did not directly predict specific internalizing and externalizing problems, but the associations were explained by maternal reflective capacity in opposite directions. More years in education predicted higher reflective capacity, which in turn predicted reductions in internalizing problems but increases in externalizing problems. These findings highlight the common (direct) and specific (indirect) pathways linking SES and SLEs to mental health outcomes, which appear to be mediated by parent socialization processes.

Emotion socialization models suggest that increased stress reduces parents' capacity to contain their own and their children's distressing emotions, which influences and is influenced by children's difficulty in managing emotions (Eisenberg et al., 1998; Eisenberg, 2020). Fonagy et al. (2021) extended emotion

socialization models to explain how economic hardship influences human development from a socio-cultural learning perspective. SLEs resulting from economic deprivation communicate to the developing child that the environment is not a safe and reliable source for social learning. Not only is the child's socialization to emotions compromised (in favour of managing the stress of limited resources), but so is their openness to and trust in cultural knowledge, i.e. explicitly and implicitly communicated norms, attitudes, beliefs, and knowledges about the self and world that are passed down the generations. Epistemic mistrust isolates a child's mind from the social world, which is the gateway to learning about the human world and all its complexities (Fonagy, Luyten, & Allison, 2015). If we are not recognized as thinking and feeling agents, or if this process is perturbed by economic adversity and inequality, then we are at risk of turning away from the social world and being left alone with our dysregulated experience (Fonagy, Luyten, Allison, & Campbell, 2019). This not only manifests as emotional and behavioural difficulties (i.e. general psychopathology; Fonagy, Luyten, Allison, & Campbell, 2017), but also difficulties in learning at school, in negotiating peer relationships, and generally fitting into the dominant system of culture.

Experimental studies suggest that participants from lower SES backgrounds tend to act in the interest of others, whereas participants from higher SES backgrounds tend to act in the interest of themselves (Rucker & Galinsky, 2017). This follows the agentic-communal model of power, whereby communities from disadvantaged backgrounds with less social power are thought to lead social lives that are more oriented towards others (e.g., promoting a shared sense of belonging and responsibility) rather than oriented towards the self (e.g., focused on personal gain and self-development; Rucker, Galinsky, & Dubois, 2017). An other-oriented

focus might result in a greater sense of togetherness and communal resilience (Liu et al., 2017; Orthner, Jones-Sanpai, & Williamson, 2004), but could also place individuals from low SES backgrounds at greater risk if their social environment is hostile and discriminatory (Fonagy et al., 2021).

The idea that an other-oriented focus poses both risk and resilience was illustrated by Milan and Wortel (2015), who found that greater sense of family obligation in adolescents from low SES backgrounds increased the positive relationship between SLEs and depression and PTSD symptoms, but reduced the positive relationship between engagement in risky behaviours and their peer group's acceptance of risky behaviours. Beliefs in the importance of family are also linked to a stronger relationship between negative romantic experiences and depression and anxiety symptoms in Mexican adolescents (Reid, Halgunseth, Espinosa-Hernandez, & Vasilenko, 2018), but a weaker relationship between cultural stressors/perceived discrimination and depressive symptoms in Latinx college studies (Corona et al., 2017). These studies highlight both the protective and exposing nature of an other-oriented focus, as indexed by family obligation, but are limited to small samples of specific sociodemographic groups.

The current study had three aims: (1) to replicate a bifactor structure of psychopathology in a large, community-based sample of older children; (2) to examine the associations between SES and general and specific psychopathology factors, and the mediating role of SLEs; and (3) to examine the role of other-orientedness in moderating the link between SES and psychopathology factors via SLEs. I predicted that a bifactor model, with a *p* factor and specific internalizing and externalizing factors would be most appropriate in summarizing the pattern of correlations among symptoms. Furthermore, I predicted that SES would modestly

and negatively predict psychopathology factors, particularly the *p* factor, and SLEs would explain a large proportion of this association. Finally, other-orientedness in the form of family obligation would predict lower levels of *p* overall, but would strengthen the positive mediating effect of SLEs on *p* factor scores, demonstrating both the protective and harmful nature of other-orientedness, respectively. I did not make specific predictions about the moderating role of family obligation on the associations between SLEs and specific psychopathology factors, as it is unclear whether there will be any remaining associations once *p* has been included in the model.

2.3 Methods

2.3.1 Participants

I analyzed data from Release 3.0 (two-year follow-up in 2020) of the Adolescent Brain Cognitive Development (ABCD) study, a large-scale longitudinal study investigating the psychological, cognitive, social, and neurobiological predictors of mental and physical health outcomes (Barch et al., 2018). I also analyzed data from Release 2.0 (one-year follow-up in 2019) in sensitivity analyses. A community sample of 11,733 children and families were self-selected from 21 nationally distributed research sites in the United States. Sites were made up of registered universities with dedicated research teams following the same assessment protocol overseen by a site quality control monitor (a list of research sites can be found here: https://abcdstudy.org/study-sites/). Children were recruited through probability sampling of mostly urban schools within each site (Garavan et al., 2018). Within each site, demographics such as age, gender, race, and

socioeconomic status were sampled, and in some cases oversampled, to reflect the US population.

Demographics for the 10,173 children who had data across all measures used can be found in Table 2.1. Most children were from White/Caucasian, upper/upper-middle class backgrounds. There was a relatively even split between children who identified as male and female. Most primary caregivers were middle-aged mothers who had graduated from college, with over half in full-time work. Indicators of poverty were generally low and neighbourhood resources were high.

Table 2.1

Socio-demographic characteristics of the ABCD sample at Release 3.0 (N = 10,173)

Sample Characteristic	*M* or *N*	*SD* or %
Child age (9-14)	11	1
Child gender		
Male	4,789	47%
Female	5,269	52%
Trans/other	50	0.50%
Missing	65	0.63%
Child's racial background		
White Caucasian	7433	73%
Black/Black African	1578	16%
Asian (Indian)	210	2%
Asian (South-East)	257	3%
Pacific Islander/Hawaii	14	0.14%
Hispanic	541	5%
Missing	140	1%
Primary caregiver age (24-81)	41	7
Primary caregiver gender		
Male	8945	88%
Female	1193	11%
Trans/other	22	0.2%
Missing	13	0.1%
Family household income		
<$5,000-$24,999	1180	12%
$25,000-$49,999	1301	13%
$50,000-$74,999	1241	12%
$75,000-$99,999	1373	13%
$100,000+	4348	43%
Missing	730	7%
Highest level of education (primary caregiver)		
No school/schooling up to Grade 12	535	5%
High school graduate or equivalent	949	9%
Some college	1634	16%
College graduate or equivalent	4337	43%
Postgraduate (Masters, Professional/Doctoral School)	2683	26%
Missing	35	0.3%
Employment status (primary caregiver)		
Full-time work	5472	54%
Part-time work	1908	19%
Not working (unable to work/not looking for work)	2556	25%
Missing	237	2%
Single-parent household	1929	19%
Number of household residents (2-16)	5	2
Total poverty markers (0-7)	0.4	0.9
Child Opportunity Index (z scores)		
Education (-.17-.23)	.02	.08

Health and Environment (-.15-.11)	.03	.04
Social and Economic (-.77-.39)	.06	.20

Note. *M* = mean; *N* = sample size; *SD* = standard deviation.

2.3.2 Measures

Psychopathology. Child psychopathology was assessed using the child-reported Brief Problem Monitor (BPM; Achenbach, McConaughy, Ivanova, & Rescorla, 2011) for ages 6-18. The BPM includes 19 questions about internalizing problems (e.g., anxiety and mood symptoms), attention problems, and externalizing problems (e.g., oppositional and destructive behaviours). Items are rated on a three-point scale (0 = Not True, 1 = Somewhat True, and 2 = Very True). I also analyzed child-reported BPM items from Release 2.0 and parent-reported BPM items from Release 3.0 in sensitivity analyses. The BPM is based on the Child Behaviour Checklist (CBCL; Achenbach & Rescorla, 2001), a widely used measure of child psychopathology, and shows good internal consistency and convergent validity with the CBCL (Pedersen et al., 2021; Piper, Gray, Raber, & Birkett, 2014).

Socioeconomic Status (SES). I developed a composite measure of SES made up of household family income, primary caregiver's highest level of education, primary caregiver's employment status, single-parent household status, number of residents living the child's household, and poverty score (e.g., the total number of poverty indicators reported, such as not being able to pay for bills; see Table 2.1 for covariate levels). I also included three indicators from the Child Opportunity Index (COI) 2.0, a set of measures reflecting US neighbourhood resources in three domains: education (e.g., access to education centers, elementary and higher education attainment rates), health and environment (e.g., access to healthy food and green spaces), and social and economic factors (e.g., economic opportunities,

single-parent households; Noelke et al., 2020). For each family's neighbourhood, metrics are collected from public sources and census data and combined into standardized scores for each domain, weighted by how strongly they predict health and economic outcomes.

Stressful Life Events. Children rated their life-time exposure (yes/no) to 25 stressful life events (SLEs) listed by Tiet et al. (1998). I analyzed the 23 life events that overlapped with the adverse childhood experiences (ACEs) documented by Felitti et al. (1998), e.g., parent mental health difficulties or witnessing domestic violence, as well 'second-generation' ACEs, e.g., parents separating or being a victim of a crime (Mersky, Janczewski, & Topitzes, 2017; see Table 2.2 for list of SLEs). I also analyzed the same 23 SLEs reported by parents in a sensitivity analysis.

Family Obligation. Children rated 16 items from three subscales of the Mexican-American Cultural Values Scale (MACVS; Knight et al., 2010): Familism-Support (e.g., maintaining close relationships with family members), Familism-Obligation (e.g., prioritizing the family's needs), and Familism-Referents (e.g., prioritizing the family's views in decision-making). Items were rated on a five-point scale (0 = Not at All; 1 = A Little; 2 = Somewhat; 3 = Very Much; 4 = Completely). The MACVS item pool was developed from focus groups with Mexican families, but family-centered values are found across diverse ethnic groups (Schwartz, 2007). The familism subscales show acceptable-to-high internal consistency and converge across family members (Knight et al., 2010; Knight et al., 2011).

2.3.3 Procedure

Children and parents (88% mothers) attended their designated research site annually to complete a battery of self-report measures, behavioural tasks, and

biological measures (neuroimaging occurred bi-annually; Karcher & Barch, 2021). Each site used the same assessment protocol overseen by a site quality control monitor to ensure adherence (Auchter et al., 2018).

2.3.4 Statistical Analysis

Factor Analysis. To address our first aim (replicating a bifactor structure of psychopathology), I estimated a bifactor model using the child-reported BPM items. The bifactor model included a general (*p*) factor with direct loadings from all items, and specific factors for each subscale (internalizing, externalizing, and attention problems). Covariances between the general factor and specific factors, and among the specific factors, were fixed at zero. I compared the bifactor model to a correlated factors model with covarying internalizing, externalizing, and attention factors, and a single factor model with one factor upon which all items loaded. For our second and third aims (examining the mediating role of SLEs and moderating role of family obligation in the relationship between SES and psychopathology factors), I estimated single factors for the SES indicators, SLEs indicators, and MACVS items).

I first ran the models using the Weighted Least Squares Means and Variances Weighted estimator to estimate measures of global fit suitable for non-normal indicators (Li, 2016). I defined acceptable and excellent incremental fit respectively as Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values ≥ 0.90 and ≥ 0.95, and acceptable and excellent absolute fit respectively defined as Root Mean Square Error of Approximation (RMSEA) values ≤ 0.08 and ≤ 0.06 (Hu & Bentler, 1999). All factor loadings reported are from models estimated using the WLSMV estimator.

I then re-ran the models using the Robust Maximum-Likelihood (MLR) estimator, which is also suitable for non-normal indicators (Li, 2016). MLR provides information criteria (e.g., AIC, BIC, and sample size adjusted BIC or BIC*n*) which are useful for comparing non-nested models like the bifactor model, correlated factors model, and single-factor model. Information criteria also offer more stringent penalties for model complexity based on the number of freely estimated parameters compared to other fit indices like TLI (Gignac & Watkins, 2013). When comparing models, I always took the difference between the baseline model and competing model (e.g., $AIC_{baseline} - AIC_{candidate}$). A difference of ≤ 2, 2-7, 7-10, and > 10 suggested little, some, strong, and very strong evidence favouring the competing model, respectively (Fabozzi, Focardi, Rachev, & Arshanapalli, 2014).

As the models estimated with MLR will be subtly different to the models estimated with WLSMV, I also compared models for their global fit statistics estimated with WLSMV. Nonetheless, I urge the reader to treat these comparisons with caution, as the models compared do not fulfil the criteria of being nested–these comparisons are more of a spot-check to ensure consistency with comparisons using information criteria, which are suitable to the current models. To this end, increases in CFI and TLI values greater than 0.01, and decreases in RMSEA greater than 0.015, between the more and less restricted models suggest an improvement in fit (Cheung & Rensvold, 2002).

Lastly, I evaluated the reliability of latent factors and observed factor scores using model-based reliability indices, with acceptable reliability indicated by Explained Common Variance (ECV) values ≥ .7 (bifactor model only), Omega (ω) values ≥ .8, Omega Hierarchical (ω_H) values ≥ .8 (bifactor model only), Construct

Reliability (*H*) values ≥ .7, and Factor Determinacy (FD) values ≥ .9 (Dueber, 2017; Rodriguez, Reise, & Haviland, 2016).

Moderated Mediation Model. I used a second-stage moderated mediation model (Preacher, Rucker, & Hayes, 2007) to investigate our second aim (how SLEs mediate the link between SES and psychopathology factors) and third aim (how family obligation moderates the indirect effect of SLEs). Figure 2.1 shows the conceptual and statistical diagrams for a second-stage moderated mediation model. As shown in the conceptual diagram, the moderator (*W*) moderates the mediator's (*M*) prediction of the outcome (*Y*; i.e. moderation at the second stage of mediation) rather than the exposure variable's (*X*) prediction of the mediator (i.e. moderation at the first stage of mediation). Statistically, mediation involves estimating a direct path (c') from *X* (e.g., SES) to *Y* (e.g., general and specific psychopathology factors), as well as indirect paths from *X* to *M* (e.g., SLEs; a_1), and from *M* to *Y* variable (b_1). The indirect effect[5] of SES on psychopathology via SLEs is estimated through the product of the a_1 and b_1 paths, while the total effect is the sum of the indirect effect (a_1b_1) and direct effect (c').

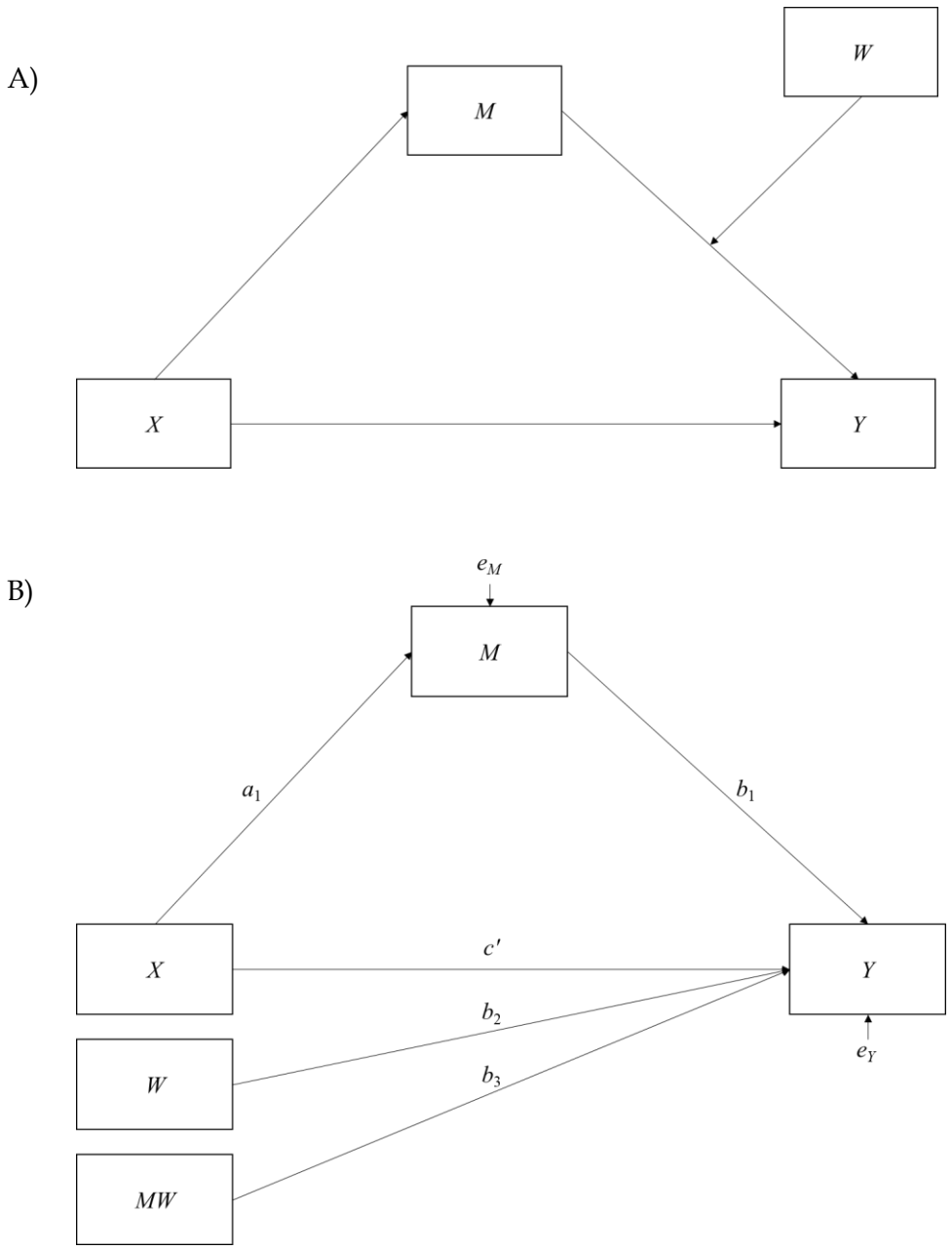
Moderation of the mediator path is estimated by direct paths from *W* (e.g., family obligation; b_2) to *Y*, and the product of *W* and *M* (*MW*; b_3) to *Y*. Moderation of the indirect effect[6] is estimated by the index of moderated mediation, which is the product of *X*'s prediction of *M* and *MW*'s prediction of *Y* (i.e. a_1b_3). Evidence for moderated mediation is apparent if bootstrapped confidence intervals for the index

[5]In using the terms direct and indirect "effects", I do not suggest that these associations are causal, but are maintaining consistency with the nomenclature of mediator analysis.
[6]The indirect effect (*IE*) is therefore equal to $IE = a_1b_1 + a_1b_3W$, which is equal to a_1b_1 when *W* is zero.

of moderated mediation do not include zero (Hayes, 2015). I used 10,000 bootstrapped samples.

Figure 2.1

Conceptual (A) and statistical (B) representations of a second-stage moderated mediation model



Note. *M* = mediator variable; *MW* = interaction term between moderator and mediator variables; *W* = moderator variable; *X* = exposure variable; *Y* = outcome variable.

I simultaneously included the *p* factor and specific internalizing, externalizing, and attention factor scores as uncorrelated outcomes to maintain the orthogonality constraint from the bifactor model. Models were estimated using robust maximum likelihood. I used observed factor scores rather than latent factors as the latter was computationally unfeasible with more than seven integration points. All analyses were conducted in Mplus 8.0 (Muthén & Muthén, 2017).

Sensitivity Analyses. I ran several sensitivity analyses to ensure the robustness of our findings. First, child age, gender (male vs. female; male vs. trans/other), and ethnicity (White vs. Black; White vs. Asian; White vs. Native American; White vs. Hispanic) were included as covariates predicting SES, SLEs, family obligation**,** and psychopathology factor scores in the moderated mediation model. Second, SES, SLEs, family obligation**,** and psychopathology factor scores were regressed on general and specific psychopathology factors estimated one year before (Release 2.0/one-year follow up) to control for the impact of pre-existing mental health difficulties driving SES's predictions (i.e. social selection). Third, I re-estimated the moderated mediation models after removing SLE indicators that overlapped with SES. Fourth, I included SES indicators as separate but correlated predictors in the moderated mediation model, due to claims that SES composites lack precision (Bradley & Corwyn, 2002). Fifth, I replicated the moderated mediation model using the correlated factor model rather than the bifactor model to determine the impact of accounting for the shared variance with the *p* factor. Sixth, I supplemented child-reported items with parent-reported items in our initial factor analytic models to estimate cross-informant factors. Finally, I also ran models using factor scores estimated from parent-reported factors only.

2.4 Results

2.4.1 Factor Analyses

Psychopathology. A bifactor model of child-reported BPM items with a *p* factor and specific internalizing, externalizing, and attention factors showed an excellent fit to the data (see Table 2.3). The *p* factor explained most of the modelled variance in item responses (ECV = .68), and most of the unit-weighted variance associated with raw total scores (ω_H = .83) and subscale scores (ω_{HS} = .15-.43). Only the *p* factor (*H* = .92) and specific internalizing factor (*H* = .71) were represented reliably by their indicators; half of externalizing and attention items loaded more strongly or exclusively onto the *p* factor compared to their respective specific factors (see Table 2.4). Only factor scores for the *p* factor met threshold for reliability (*FD* = .93).

The correlated factors model with internalizing, externalizing and attention factors also showed an excellent fit to the data (see Table 2.3). The correlated factors showed strong positive loadings, were well represented by their indicators, and showed strong positive loadings (see Table 2.4). They also showed strong correlations, suggesting the presence of common factor. By contrast, the single factor did not show acceptable fit (see Table 2.3) but had strong positive loadings (see Table 2.4). The bifactor model fit better than the correlated factors model (ΔAIC = 1055.51, ΔBIC = 856.39, ΔBICn = 907.25; ΔCFI = .01; ΔTLI = .01; ΔRMSEA = .006) and single factor model (ΔAIC = 6994.35, ΔBIC = 6773.55, ΔBICn = 6833.94; ΔCFI = .11; ΔTLI = .12; ΔRMSEA = .048). Furthermore, the correlated factors model fit better than the single factor model (ΔAIC = 5938.85, ΔBIC = 5917.17, ΔBICn = 5926.70; ΔCFI = .10; ΔTLI = .11; ΔRMSEA = .042).

SLEs. A single factor model including all 23 SLEs did not converge because children did not report exposure to the following SLEs: 'being placed in foster care', 'family became homeless', 'parent/caregiver was hospitalized', 'school locked down due to violence', 'saw/heard a shooting', 'know of someone who attempted suicide', and 'parent/caregiver was deported'. After removing these SLEs, a single factor model showed excellent absolute fit but fell below acceptable incremental fit (see Table 2.3). SLEs loaded positively and moderately onto the SLE factor, which showed reliable factor scores, total score reliability, and construct reliability (see Table 2.2).

SES. A single factor model with all nine SES indicators did not show acceptable absolute or incremental fit (see Table 2.3). Furthermore, employment status (λ = -.22) and number of residents (λ = .04) showed weak loadings. After removing these two items, model fit improved but fell under acceptable criteria. Nonetheless, factor score, total score, and construct reliability for SES factor scores was high (see Table 2.2).

Family Obligation. A single 'family obligation' factor model showed acceptable incremental fit but not absolute fit (see Table 3). All MASC items loaded positively and strongly onto the family obligation factor, which showed high levels of factor score, total-score, and construct reliability (see Table 2).

Table 2.2

Standardized factor loadings for the final single factor solutions of the Stressful Life Events (SLE) checklist, Socioeconomic Status (SES) indicators, and Mexican-American Cultural Values Scale (MACVS)

SLE Checklist	λ	SES Indictors	λ	MACVS	λ
Negative change in parents' financial situation	0.61	Household income	0.89	Children should be taught to always be good because they represent the family.	0.79
Family member had a mental/emotional problem	0.58	COI Social and Economic	0.78	It is important to work hard and do one's best because this work reflects on the family.	0.79
Family member had a drug/alcohol problem	0.55	COI Education	0.68	It is always important to be united as a family	0.78
Parents argued more than before	0.55	Highest level of education	0.66	A person should always think about their family when making important decisions.	0.78
I was a victim of a crime/assault/violence	0.51	COI Health and Education	0.65	It is important for family members to show their love and affection to one another.	0.75
I got seriously sick	0.48	Single-parent household	0.53	It is important to have close relationships with aunts/uncles, grandparents, and cousins.	0.74
Close friend died	0.48	Poverty Score	0.36	Holidays and celebrations are important because the whole family comes together.	0.72
Parent lost their job	0.47			Children should always do things to make their parents happy.	0.71
Parent was away from home more	0.46			Family provides a sense of security because they will always be there for you.	0.71
Family member was seriously injured	0.46			If a relative is having a hard time financially, one should help them out if possible.	0.71
Lost a close friend	0.45			Older kids should take care of and be role models for their younger brothers and sisters.	0.71

Close friend seriously sick/injured	0.45		Children should be taught that it is their duty to care for their parents when their parents get old.	0.7
I got seriously injured	0.45		When it comes to important decisions, the family should ask for advice from close relatives.	0.7
Witnessed crime/accident	0.41		A person should share their home with relatives if they need a place to stay	0.69
Parents separated/divorced	0.41		Parents should teach their children that the family always comes first.	0.66
Family member died	0.24		Parents should be willing to make great sacrifices to make sure their children have a better life.	0.65
M	0.47	0.65		0.72
SD	0.08	0.17		0.04
ω	0.82	0.84		0.95
H	0.83	0.89		0.97
FD	0.91	0.94		0.95

Note. COI = Child Opportunity Index; *FD* = Factor Determinacy; *H* = Construct Reliability; ω = Omega; λ = factor loading.

Table 2.3

Model Fit Statistics for Factor Models of the Brief Problem Monitor (BPM), Stressful Life Events (SLE) Checklist, Socioeconomic Status (SES) Indicators, and Mexican-American Cultural Values Scale (MACVS) items

Measure/Model	χ^2 (*df*)	RMSEA	CFI	TLI	AIC	BIC	BIC*n*
BPM							
Bifactor	2,099.46 (133)	.038 (.037-.04)	.98	.97	226,194.60	226,827.32	226,585.79
Correlated Factors	3,094.32 (149)	.044 (.043-.045)	.97	.96	227,250.11	227,683.70	227,493.03
Single Factor	11,566.68 (152)	.086 (.085-.087)	.87	.85	233,188.95	233,600.87	233,419.73
SLE							
Single Factor	15,20.94 (104)	.037 (.035-.038)	.86	.85	124,208.01	124,439.27	124,337.58
SES							
All items	5,828.28 (27)	.15 (.14-.15)	.77	.70	70,621.70	70,838.53	70,743.19
Removed employment and resident number	4,521.20 (29)	.12 (.12-.13)	.82	.78	70,788.91	70,991.28	70,902.30
MACVS							
Single Factor	8,240.57 (104)	.088 (.086-.089)	.95	.95	336,184.17	336,762.15	336,507.93

Note. AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; BIC*n* = sample-sized corrected Bayesian Information Criteria; CFI = Comparative Fit Index; *df* = Degrees of Freedom; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation; χ^2 = Chi-square.

Table 2.4

Standardized factor loadings for the bifactor model, correlated factors model, and single factor model of the Brief Problem Monitor child-reported items

	Model							
	Bifactor				Correlated Factors			Single Factor
Item	*p*	Ext	Int	Att	Ext	Int	Att	*p*
Acts young	0.54			0.02			0.54	0.50
Argues	0.61	0.46			0.73			0.62
Can't finish things	0.58			0.19			0.63	0.58
Can't concentrate	0.67			0.61			0.84	0.78
Can't sit still	0.55			0.43			0.68	0.63
Destroys things	0.66	0.22			0.66			0.58
Disobedient (home)	0.61	0.24			0.67			0.58
Disobedient (school)	0.66	0.14			0.65			0.57
Worthless	0.64		0.49			0.84		0.74
Impulsive	0.74			0.00			0.71	0.65
Fearful	0.54		0.61			0.80		0.69
Guilty	0.58		0.51			0.76		0.67
Embarrassed	0.50		0.48			0.72		0.62
Distracted	0.68			0.50			0.84	0.78
Stubborn	0.52	0.28			0.63			0.54
Temper	0.60	0.44			0.73			0.62
Threatens	0.65	0.39			0.70			0.60
Unhappy	0.56		0.47			0.76		0.67
Worries	0.50		0.61			0.76		0.65
M	0.60	0.31	0.53	0.29	0.68	0.77	0.71	0.63
SD	0.07	0.12	0.07	0.26	0.04	0.04	0.12	0.08
ECV	0.68	0.07	0.17	0.08				
ω	0.94	0.87	0.90	0.86	0.86	0.90	0.86	0.93
ω_H	0.83	0.18	0.43	0.15				
Rel. ω	0.88	0.20	0.48	0.18				
H	0.92	0.47	0.71	0.54	0.86	0.90	0.88	0.93
FD	0.93	0.69	0.85	0.80	0.93	0.95	0.94	0.97

	Ext	Int
Ext		
Int	.58	
Att	.74	.63

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; ω = Omega; ω_H = Omega Hierarchical.

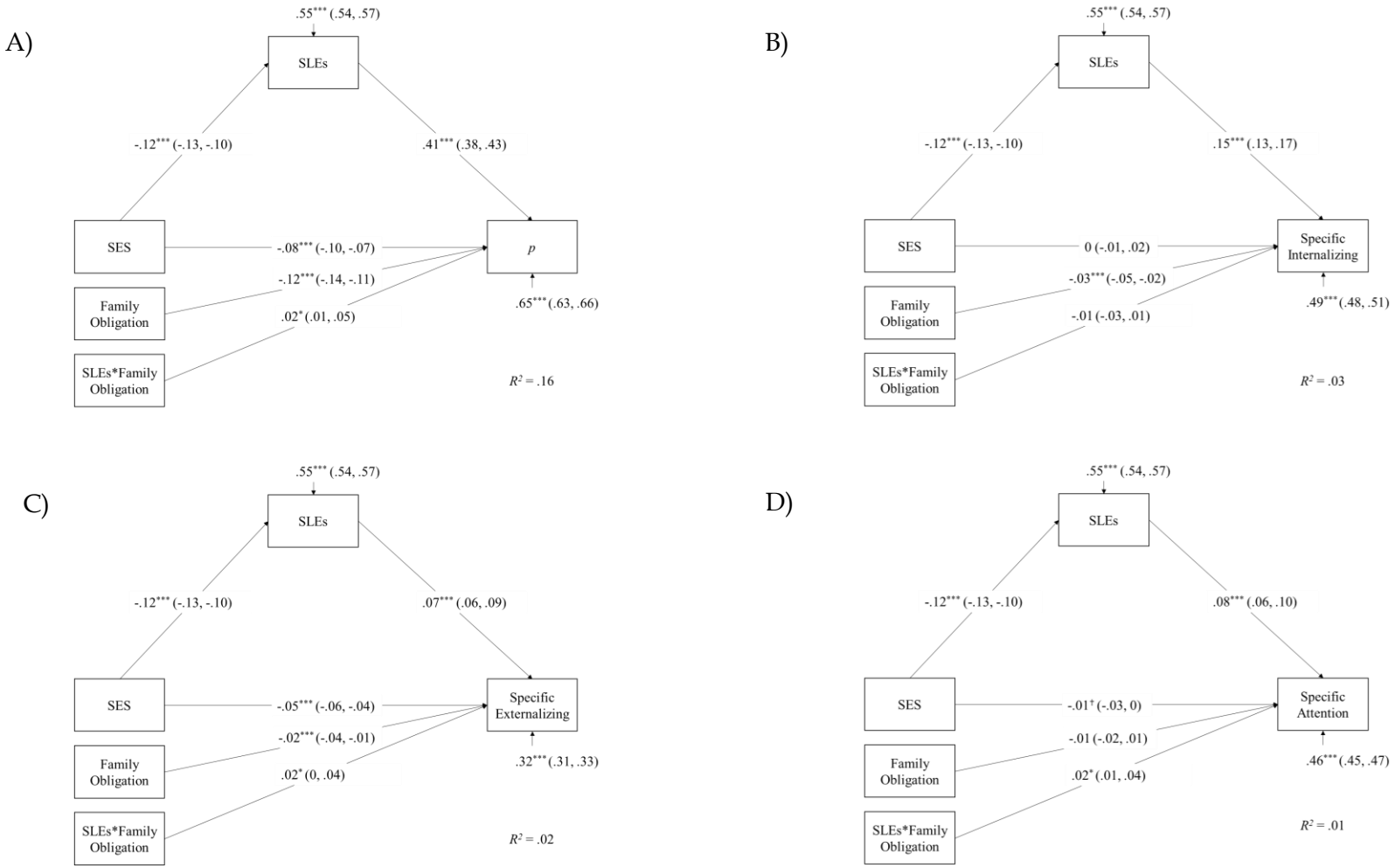
2.4.2 Moderated Mediation

Figure 2.2 shows the moderated mediation model, with SES factor scores predicting the *p* factor scores and specific internalizing, externalizing, and attention factor scores directly and indirectly via lifetime SLE factor scores. The indirect effect of SLEs on the *p* factor and specific factors is moderated by family obligation factor scores.

Regarding the direct paths (e.g., c'), higher SES factor scores weakly predicted lower *p* factor scores, lower specific externalizing scores, and marginally lower specific attention scores, but not specific internalizing scores. As for the indirect paths, higher SES factor scores weakly but significantly predicted lower SLE factor scores (a_1), and higher SLE factor scores moderately predicted higher *p* factor scores and higher specific internalizing, externalizing, and attention factor scores (b_1). The indirect effect of SES via SLEs (e.g., a_1b_1) was significant and negative for *p* factor and specific externalizing and attention scores, whereby SLEs suppressed the negative effect of SES on psychopathology scores. Put differently, higher SES predicted a reduction in psychopathology scores through a reduction in SLEs (and lower SES predicted higher psychopathology scores through higher SLEs). For specific internalizing factor scores, SLEs reduced the small but positive effect of SES on internalizing, whereby higher SES and SLEs predicted higher internalizing scores, but higher SES predicted lower SLEs which weakened its prediction of internalizing scores. Indirect effects were generally small to moderate relative to the total effect (see Table 2.5).

Figure 2.2

Moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the p factor (A), specific internalizing factor (B), specific externalizing factor (C), and specific attention factor (D)



***$p \leq .001$; **$p \leq .01$; *$p \leq .05$; †$p \leq .10$

As for the moderator paths, starting with the b_2 paths, higher family obligation factor scores significantly but weakly predicted lower *p* factor scores and lower specific internalizing and externalizing factor scores, but not attention scores (see Figure 2.2). The interaction between family obligation and SLE factor scores (e.g., b_3) was significant and positive for the *p* factor and specific externalizing and attention factor scores, but not specific internalizing factor scores, such that SLE's positive prediction of psychopathology scores was stronger at higher family obligation scores. The moderated mediation effect, estimated by the index of moderated mediation (e.g., a_1b_3), was significant and negative for *p* factor scores and specific externalizing and attention scores, but not internalizing scores (see Table 2.5). Family obligation suppressed the negative indirect effect of SES via SLEs on psychopathology scores, such that the indirect effect became more negative (and hence stronger) at higher levels of family obligation. In other words, the way in which higher SES predicted a reduction in psychopathology scores through a reduction in SLEs was stronger at higher levels of family obligation, and by the same token, lower SES predicted higher psychopathology scores through higher SLEs most strongly at higher levels of family obligation (see IMMs in Table 2.5).

Table 2.5

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status via Stressful Life Events

	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
P	-.13***	-.15, -.11	-.05***	-.06, -.04	36%	-.003*	-.005, -.001
Internalizing	-.02*	-.03, -.01	-.02***	-.02, -.01	88%	.001	-.001, .003
Externalizing	-.06***	-.07, -.05	-.01***	-.01, -.01	14%	-.002*	-.004, -.001
Attention	-.02**	-.04, -.01	-.01***	-.01, -.01	41%	-.003*	-.005, -.001

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.

2.4.3 Sensitivity Analyses

A moderated mediation model controlling for age, gender, and ethnicity produced similar findings to the moderated mediation model without covariates (see Appendix 2, Tables A2.1 and A2.2, and Figure A2.1). Moreover, a moderated mediation model controlling for *p* factor and specific internalizing, externalizing, and attention factor scores estimated one year earlier showed similar but weaker path coefficients, with some changes to the specific internalizing and attention paths (see Tables A2.3-A2.5 and Figure A2.2). SLE factor scores estimated from a factor without SLE indicators that overlapped with SES produced similar path coefficients (see Table A2.6 and Figure A2.3).

When SES indicators were included as separate but correlated predictors in the moderated mediation model, higher household income and lower poverty scores indirectly predicted lower *p* factor and specific attention scores via lower SLEs, particularly at higher levels of family obligation (see Tables A2.7-A2.8). Moreover, the presence of a partner in the family household predicted higher *p* factor (marginal) and specific internalizing scores, but was suppressed by higher SLE scores, particularly at higher levels of family obligation (see Tables A2.7-A2.8).

In a moderated mediation model with correlated internalizing, externalizing, and attention factor scores estimated from a correlated factors model, direct and indirect paths related to externalizing and attention scores mirrored those of the *p* factor in direction and strength (see Figure A2.4). Moreover, SES positively predicted internalizing scores in the correlated factors version, but not in the bifactor version. The indirect effect of SES via SLEs was moderated by family

obligation for externalizing and attention scores, but not for internalizing scores, like in the bifactor version (see Table A2.9).

Finally, child- and parent-reported BPM, SLE, and MACVS items did not converge when estimating cross-informant general and specific psychopathology factors, an SLE factor, and a family obligation factor, respectively (see Tables A2.11-A2.14). A moderated mediation model using only parent-reported items showed similar results to the child-reported version, but SES now significantly and positively predicted specific internalizing and attention scores, while family obligation scores switched from negatively to positively predicting specific externalizing scores (see Tables A2.16 and Figure A2.5).

2.5 Discussion

In this chapter, I investigated the pathways linking SES to common and specific features of child mental health problems. I first replicated a bifactor structure of psychopathology, with a general *p* factor and specific internalizing, externalizing, and attention factors. I then ran a moderated mediation model in which higher SES predicted a reduction in *p* factor and specific externalizing and attention scores through a reduction in SLEs. The indirect effect of SLEs in explaining the link between SES and the *p* factor and specific externalizing and attention problems was stronger at higher levels of family obligation.

2.5.1 Aim 1. Replicating a Bifactor Psychopathology Structure in Older Children

As predicted, a bifactor model with a general *p* factor and specific internalizing, externalizing, and attention factors suited both the child-reported and

parent-reported psychopathology data. Previous studies have also favoured a bifactor model when summarizing psychopathology data in older children (Afzali, Sunderland, Carragher, & Conrod, 2017; Carragher et al., 2016), including data from prior waves of the ABCD dataset (Clark et al., 2021; Moore et al., 2020). We cannot conclude that the latent structure of psychopathology is bifactorial just because our bifactor model showed a slight advantage in model fit compared to the correlated factors and single factor models (Greene et al., 2019; Sellbom & Telegen, 2019). However, I chose the bifactor model as my intention was to tease apart the shared and unique variance in symptom responses.

Some have interpreted the *p* factor in childhood as reflecting a dysregulation profile, i.e. widespread difficulties in regulating attention, affect, and behaviour (Brandes, Herzhoff, Smack, & Tackett, 2019; Deutz et al., 2020; Haltigan et al., 2018). Our bifactor solution showed the strong presence of a *p* factor, but it was characterized by difficulties in attention and behaviour more so than affect. This might be because persistent externalizing problems are typical of the dysregulation profile in childhood (Basten et al., 2013). Alternatively, externalizing items may have represented the common variance more reliably because they are more frequent in the BPM. These explanations are probably related (e.g., externalizing problems were better represented in the item pool because they are more representative of the dysregulation profile in childhood). However, I am mindful of having estimated one version of the *p* factor among a universe of *p* factors that are affected by the methods used to measure them, including the sample characteristics and measure composition (Constantinou & Fonagy, under review).

2.5.2 Aim 2. The relationship between SES and psychopathology factors and the mediating role of SLEs.

Using the bifactor model, I was able to tease apart the associations between SES and the common and specific features of child mental health difficulties. SES was most predictive of the *p* factor and specific externalizing problems, which coincides with prior studies showing that SES negatively predicts both internalizing and externalizing problems (i.e. domain-general effects), particularly externalizing problems (i.e. domain-specific effects; Peverill et al., 2021), and associations between SES and externalizing problems remain even after accounting for the common variance via the *p* factor (Lahey et al., 2012; Liu et al., 2017; Patalay et al., 2015). By contrast, associations between SES and specific internalizing and attention problems were negligible, despite being present in the correlated factors model. This implies that the relationship between SES and the correlated internalizing and attention factors is driven by shared features that are controlled for by the *p* factor in the bifactor model.

Like past studies, the relationship between SES and the *p* factor and specific externalizing factor was partially mediated by SLEs (Amone-P'olak et al., 2009; Kang et al., 2011; Ziebold et al., 2021). However, the *p* factor and specific externalizing factor differed in the extent that SLEs mediated these associations. More than a third of SES's total effect on the *p* factor was explained by SLEs, suggesting some importance of the indirect (proximal) pathway in addition to the direct (distal) pathway to child mental health difficulties (Peverill et al., 2021; Reiss, 2013). In contrast, 14% of SES's total effect on specific externalizing problems was explained by SLEs, emphasizing the direct importance of socioeconomic resources to behavioural difficulties. Differences in the strength of SLEs' indirect effect may

reflect different causal pathways: the *p* factor may captures difficulties related to relational stress, whereas the specific externalizing factor may capture difficulties related to adaptive pressures towards antisocial behaviour (Liu et al., 2017; Snyder, Young, & Hankin, 2019).

2.5.3 Aim 3. The role of family obligation in moderating the link between SES and psychopathology factors via SLEs.

I found that family obligation, i.e. prioritizing the needs and views of the family over one's own, served both protective and harmful roles. On the one hand, higher levels of family obligation directly predicted lower levels of the *p* factor and specific internalizing and externalizing problems. This finding is consistent with studies showing that family-centered values (Corona et al., 2017; Milan & Wortel, 2015), or community-centered values (Liu et al., 2017), reduce the risk of child mental health difficulties. On the other hand, the indirect effect of SLEs on the *p* factor and specific externalizing and attention factors scores was strongest at higher levels of family obligation. This finding replicates studies showing that higher family obligation in children and adolescents from low socioeconomic or ethnic minority backgrounds predicts a stronger relationship between SLEs and depression and anxiety symptoms (Milan & Wortel, 2015; Reid et al., 2018).

The dual role of family obligation can be understood from a socio-cultural learning perspective (Fonagy et al., 2021). Prioritizing the family's needs and views reflects an other-oriented focus that would allow families and groups more broadly to act and think together as a collective (Tuomela & Tuomela, 2005). When a child's mind is recognized as part of a psychological collective, it fosters a special type of learning that marks the personal and communal relevance of what is being shared,

i.e. cultural knowledges about the self, others, and world (Fonagy et al., 2015). When the social environment is relatively benign and attuned to the child's experience, a sense of trust in social learning develops. Epistemic trust allows a child to update their understandings about themselves and others in an everchanging world, and avoid a rigidity in beliefs that results in dysregulated affect, behaviour, and thought (i.e. general psychopathology; Fonagy et al., 2017). However, when the social environment is misattuned and features adversity (e.g., stressful life events), the child who is oriented towards others will develop an aversion towards social learning (Fonagy et al., 2021). Epistemic mistrust prevents the child from updating their understandings of themselves and others because the source of learning–the social environment–is felt to be unsafe, but at the cost of being stuck in prior ways of thinking, feeling, and acting (Fonagy et al., 2019).

It should be stressed the moderated mediation effect was small and requires replication. However, the effect did hold even after controlling for demographics, particularly racial background, since family obligation tends to be higher in ethnic minorities (Schwartz, 2007) and was so in Black and Hispanic Americans in our study. Still, the harmful nature of family obligation may not be due to social learning mechanisms as I have interpreted. Instead, children who are family-focused and have been exposed to adversity may be more prone to mental health difficulties for fear of disapproval or prioritizing their own needs.

2.5.4 Strengths and Limitations

I replicated the mediating role of SLEs and moderating role of family obligation in a large, representative sample of older children, after testing for several confounds in sensitivity analyses. Using the bifactor model, I could estimate

associations with SES that are likely to be the product of the commonalities among child mental health difficulties (e.g., SES no longer significantly predicted internalizing and attention problems once its associations with the *p* factor were accounted for). Had I just used the correlated factors model, I might have concluded that SES predicts internalizing, externalizing, and attention domains equally.

There are, however, disadvantages associated with our bifactor solution. Specific psychopathology factor scores showed poor reliability compared to recommended standards (Rodriguez et al., 2016). I therefore caution any firm conclusions from our analyses of specific factor scores, which might partially explain the lack of significant associations between SES and specific internalizing and attention factors. Furthermore, I did not validate our factors against external criteria, so cannot conclude with certainty that they reflect substantive constructs such as general dysregulation.

One may also question our use of factor analysis for analyzing SES indicators. Some claim that composite measures of SES lose important information leading to false claims about the (lack of) relationship between SES and child mental health outcomes (Bradley & Corwyn, 2002). I analyzed SES indicators as a composite measure as well as separate but parallel predictors. While I found that certain indicators when analyzed separately were more consistently associated with psychopathology factor scores than others (e.g., household income, presence of partner, socioeconomic neighbourhood opportunities), both composite and separate SES indicators predicted a similar amount of variance.

Markers of socioeconomic disadvantage tend to co-occur in reality: families with a low household income are also more likely to be single-parent households with a caregiver who has had fewer educational opportunities (Singh, 2003). The

same is true for SLEs: children with parents who have mental health difficulties are also more likely to be exposed to parental conflict, separation, and domestic violence (Lacey & Minnis, 2020). Yet, some markers of SES or SLEs may be more influential than others, such as a parent dying compared to a parent with a treatable illness. Factor models allow us to represent the overlap among indicators whilst weighting their relative importance.

Still, our findings are limited to families who fall at the upper end of the SES spectrum; this may have weakened the associations between SES and psychopathology factors. Our of SES and SLEs are also limited to our 'objective' perspective as researchers, since I analyzed exposure to SLEs that match an established list of ACEs, and markers of SES agreed by researchers, such as household status. There is a growing body of research highlighting the importance of perceived SES (Quon & McGrath, 2014) and subjective SLEs (Danese & Widom, 2020) in predicting mental health outcomes. Finally, I analyzed data collected during the Covid-19 pandemic in 2020. I cannot ignore the impact this may have had on children's mental health outcomes, particularly those from lower socioeconomic backgrounds who have been disproportionately affected by the pandemic (de Figueiredo et al., 2021).

2.5.5 Future Directions

Although SES partially predicted child mental health outcomes via SLEs, not all children with exposure to SLEs reported more difficulties. This raises the question: how do SLEs translate into an increased risk for child mental health difficulties and what may buffer this process? Wade et al. (2021) showed that parent process variables such as maternal responsiveness and reflective functioning

mediated the links between SES and the *p* factor and specific internalizing and externalizing factors. We could hypothesize that children in the ABCD dataset who were exposed to more SLEs but whose caregivers were more responsive and reflective may have shown lower psychopathology scores.

Nonetheless, we cannot assume that family processes like parental reflexivity affect all children in the same way; I found that the indirect effect of SLEs was stronger in children with a greater sense of duty towards the family. We can therefore ask what other cultural processes moderate the link between family process variables and psychopathology, including culturally-specific beliefs about mental health and help-seeking, different roles of family members in child socialization, and the child's degree of independence vs. inter-dependence (Bornstein, 2013). The sooner we move towards culturally sensitive analyses and assessments, the better equipped we will be at understanding risk and resilience, and engaging families clinically from diverse backgrounds (Sanchez et al., 2022).

2.5.6 Conclusions

In all, socioeconomic disadvantage is associated with increased domain-general and externalizing-specific difficulties in older children through increases in stressful life events. Children's obligation to the family's needs plays both protective and harmful roles, predicting lower levels of psychopathology overall, but amplifying the impact of socioeconomic disadvantage via increased stressful life events.

2.6 References

Achenbach T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Achenbach T. M., McConaughy, S, H, Ivanova, M. Y., & Rescorla, L. A. (2011). *Manual for the ASEBA Brief Problem Monitor™*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Afzali, M. H., Sunderland, M., Carragher, N., & Conrod, P. (2018). *The Structure of Psychopathology in Early Adolescence: Study of a Canadian Sample: La structure de la psychopathologie au debut de l'adolescence: etude d'un echantillon canadien*. Can J Psychiatry, *63*(4), 223-230. doi:10.1177/0706743717737032

Amone-P'Olak, K., Ormel, J., Huisman, M., Verhulst, F. C., Oldehinkel, A. J., & Burger, H. (2009). Life stressors as mediators of the relation between socioeconomic position and mental health problems in early adolescence: the TRAILS study. *J Am Acad Child Adolesc Psychiatry, 48*(10), 1031-1038. doi:10.1097/CHI.0b013e3181b39595

Auchter, A. M., Hernandez Mejia, M., Heyser, C. J., Shilling, P. D., Jernigan, T. L., Brown, S. A., . . . Dowling, G. J. (2018). A description of the ABCD organizational structure and communication framework. *Dev Cogn Neurosci, 32*, 8-15. doi:10.1016/j.dcn.2018.04.003

Barch, D. M., Albaugh, M. D., Avenevoli, S., Chang, L., Clark, D. B., Glantz, M. D., . . . Sher, K. J. (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci, 32*, 55-66. doi:10.1016/j.dcn.2017.10.010

Basten, M. M., Althoff, R. R., Tiemeier, H., Jaddoe, V. W., Hofman, A., Hudziak, J. J., . . . van der Ende, J. (2013). The dysregulation profile in young children: empirically defined classes in the Generation R study. *J Am Acad Child Adolesc Psychiatry, 52*(8), 841-850 e842. doi:10.1016/j.jaac.2013.05.007

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Belsky, D. W., Caspi, A., Arseneault, L., Corcoran, D. L., Domingue, B. W., Harris, K. M., . . . Odgers, C. L. (2019). Genetics and the geography of health, behaviour and attainment. *Nat Hum Behav, 3*(6), 576-586. doi:10.1038/s41562-019-0562-1

Blanco, C., Wall, M. M., Hoertel, N., Krueger, R. F., Liu, S.-M., Grant, B. F., & Olfson, M. (2019). Psychiatric disorders and risk for multiple adverse outcomes: a national prospective study. *Molecular Psychiatry, 26*(3), 907-916. doi:10.1038/s41380-019-0459-4

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. *Biol Psychiatry*, *88*(1), 18-27. doi:10.1016/j.biopsych.2020.01.013

Bornstein, M. H. (2013). Parenting and child mental health: a cross-cultural perspective. *World Psychiatry, 12*(3), 258-265. doi:10.1002/wps.20071

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annu Rev Psychol, 53*, 371-399. doi:10.1146/annurev.psych.53.100901.135233

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. doi:10.1177/2167702619859332

Brieant, A. E., Sisk, L. M., & Gee, D. G. (2021). Associations among negative life events, changes in cortico-limbic connectivity, and psychopathology in the ABCD Study. *Dev Cogn Neurosci, 52*, 101022. doi:10.1016/j.dcn.2021.101022

Bronfenbrenner, U., & Crouter, A. C. (1983). The evolution of environmental models in developmental research. In P. H. Mussen (Ed.), *Handbook of child psychology: Vol. 1. History, theory, and methods* (4th ed., pp. 357–414). New York: Wiley.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*, *2*(2), 119-137. doi:10.1177/2167702613497473

Carragher, N., Teesson, M., Sunderland, M., Newton, N. C., Krueger, R. F., Conrod, P. J., . . . Slade, T. (2016). The structure of adolescent psychopathology: a symptom-level analysis. *Psychol Med*, *46*(5), 981-994. doi:10.1017/S0033291715002470

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., . . . Sripada, C. (2021). The General Factor of Psychopathology in the Adolescent Brain Cognitive Development (ABCD) Study: A Comparison of Alternative Modeling Approaches. *Clin Psychol Sci*, *9*(2), 169-182. doi:10.1177/2167702620959317

Constantinou, M. P. & Fonagy, P. (under review). Evaluating Bifactor Models of Psychopathology Using Model-Based Reliability Indices. *Clinical Psychology Review*.

Corona, R., Rodriguez, V. M., McDonald, S. E., Velazquez, E., Rodriguez, A., & Fuentes, V. E. (2017). Associations between Cultural Stressors, Cultural Values, and Latina/o College Students' Mental Health. *J Youth Adolesc, 46*(1), 63-77. doi:10.1007/s10964-016-0600-5

Danese, A., & Widom, C. S. (2020). Objective and subjective experiences of child maltreatment and their relationships with psychopathology. *Nat Hum Behav, 4*(8), 811-818. doi:10.1038/s41562-020-0880-3

de Figueiredo, C. S., Sandre, P. C., Portugal, L. C. L., Mazala-de-Oliveira, T., da Silva Chagas, L., Raony, I., . . . Bomfim, P. O. (2021). COVID-19 pandemic impact on children and adolescents' mental health: Biological, environmental, and social factors. *Prog Neuropsychopharmacol Biol Psychiatry, 106*, 110171. doi:10.1016/j.pnpbp.2020.110171

Dueber, D. M. (2017). *Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models*. Retrieved from https://doi.org/10.13023/edp.tool.01

Deutz, M. H. F., Geeraerts, S. B., Belsky, J., Dekovic, 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 Hum Dev, 51*(1), 114-126. doi:10.1007/s10578-019-00916-2

Devenish, B., Hooley, M., & Mellor, D. (2017). The Pathways Between Socioeconomic Status and Adolescent Outcomes: A Systematic Review. *Am J Community Psychol, 59*(1-2), 219-238. doi:10.1002/ajcp.12115

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization Mediators of the Relation between Socioeconomic Status and Child Conduct Problems. *Child Development, 65*(2), 649. doi:10.2307/1131407

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic Deprivation and Early Childhood Development. *Child Development, 65*(2), 296. doi:10.2307/1131385

Eisenberg, N. (2020). Findings, issues, and new directions for research on emotion socialization. *Developmental Psychology*, *56*(3), 664–670. https://doi.org/10.1037/dev0000906

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental socialization of emotion. *Psychological Inquiry*, *9*(4), 241-273.

Fabozzi, F., Focardi, S., Rachev, S., & Arshanapalli, B. (2014). *The Basics of Financial Econometrics*. Hoboken, NJ: John Wiley & Sons.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. *American Journal of Preventive Medicine, 14*(4), 245-258. doi:10.1016/s0749-3797(98)00017-8

Felner, R. D., Brand, S., DuBois, D. L., Adan, A. M., Mulhall, P. F., & Evans, E. G. (1995). Socioeconomic Disadvantage, Proximal Environmental Experiences, and Socioemotional and Academic Adjustment in Early Adolescence: Investigation of a Mediated Effects Model. *Child Development, 66*(3), 774. doi:10.2307/1131950

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2021). Culture and psychopathology: An attempt at reconsidering the role of social learning. *Dev Psychopathol*, 1-16. doi:10.1017/S0954579421000092

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic Petrification and the Restoration of Epistemic Trust: A New Conceptualization of Borderline

Personality Disorder and Its Psychosocial Treatment. *J Pers Disord, 29*(5), 575-609. doi:10.1521/pedi.2015.29.5.575

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personal Disord Emot Dysregul, 4*, 9. doi:10.1186/s40479-017-0062-8

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, Epistemic Trust and the Phenomenology of Psychotherapy. *Psychopathology, 52*(2), 94-103. doi:10.1159/000501526

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. doi:10.1016/j.brat.2019.103411

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., . . . Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Dev Cogn Neurosci, 32*, 16-22. doi:10.1016/j.dcn.2018.04.004

Gignac, G. E., & Watkins, M. W. (2013). Bifactor Modeling and the Estimation of Model-Based Reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*(5), 639–662. https://doi.org/10.1080/00273171.2013.804398

González, L., Estarlich, M., Murcia, M., Barreto-Zarza, F., Santa-Marina, L., Simó, S., . . . Rebagliato, M. (2021). Poverty, social exclusion, and mental health: the role of the family context in children aged 7–11 years INMA mother-and-child cohort study. *European Child & Adolescent Psychiatry*. doi:10.1007/s00787-021-01848-w

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., . . . Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *J Abnorm Psychol*, *128*(7), 740-764. doi:10.1037/abn0000434

Haltigan, J. D., Aitken, M., Skilling, T., Henderson, J., Hawke, L., Battaglia, M., . . . Andrade, B. F. (2018). "P" and "DP:" Examining Symptom-Level Bifactor Models of Psychopathology and Dysregulation in Clinically Referred Children and Adolescents. *J Am Acad Child Adolesc Psychiatry*, *57*(6), 384-396. doi:10.1016/j.jaac.2018.03.010

Hashimoto, K., Reiss, F., Meyrose, A.-K., Otto, C., Lampert, T., Klasen, F., & Ravens-Sieberer, U. (2019). Socioeconomic status, stressful life situations and mental health problems in children and adolescents: Results of the German BELLA cohort-study. *PLoS One, 14*(3), e0213700. doi:10.1371/journal.pone.0213700

Hayes, A. F. (2015). An Index and Test of Linear Moderated Mediation. *Multivariate Behav Res, 50*(1), 1-22. doi:10.1080/00273171.2014.962683

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyland, P., Murphy, J., Shevlin, M., Bentall, R. P., Karatzias, T., Ho, G. W. K., . . . McElroy, E. (2021). On top or underneath: where does the general factor of psychopathology fit within a dimensional model of psychopathology? *Psychol Med, 51*(14), 2422-2432. doi:10.1017/S003329172000104X

Kang, E., Mellins, C. A., Dolezal, C., Elkington, K. S., & Abrams, E. J. (2011). Disadvantaged Neighborhood Influences on Depression and Anxiety in Youth with Perinatally Acquired Human Immunodeficiency Virus: How

Life Stressors Matter. *J Community Psychol, 39*(8), 956-971. doi:10.1002/jcop.20483

Karcher, N. R., & Barch, D. M. (2020). The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology, 46*(1), 131-142. doi:10.1038/s41386-020-0736-6

Knight, G. P., Gonzales, N. A., Saenz, D. S., Bonds, D. D., Germán, M., Deardorff, J., … Updegraff, K. A. (2009). The Mexican American Cultural Values Scale for Adolescents and Adults. *The Journal of Early Adolescence*, *30*(3), 444–481. doi:10.1177/0272431609338178

Knight, G. P., Berkel, C., Umaña-Taylor, A. J., Gonzales, N. A., Ettekal, I., Jaconis, M., & Boyd, B. M. (2011). The Familial Socialization of Culturally Related Values in Mexican American Families. *Journal of Marriage and Family, 73*(5), 913-925. doi:10.1111/j.1741-3737.2011.00856.x

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., . . . Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annu Rev Clin Psychol, 17*, 83-108. doi:10.1146/annurev-clinpsy-081219-093304

Lacey, R. E., & Minnis, H. (2020). Practitioner Review: Twenty years of research with adverse childhood experience scores - Advantages, disadvantages and applications to practice. *J Child Psychol Psychiatry, 61*(2), 116-130. doi:10.1111/jcpp.13135

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? *J Abnorm Psychol*, *121*(4), 971-977. doi:10.1037/a0028355

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. doi:10.3758/s13428-015-0619-7

Liu, J., Mustanski, B., Dick, D., Bolland, J., & Kertes, D. A. (2017). Risk and protective factors for comorbid internalizing and externalizing problems among economically disadvantaged African American youth. *Dev Psychopathol, 29*(3), 1043-1056. doi:10.1017/S0954579416001012

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol*, *15*, 51-69. doi:10.1146/annurev-clinpsy-050718-095522

Martin, M. J., Conger, R. D., Schofield, T. J., Dogan, S. J., Widaman, K. F., Donnellan, M. B., & Neppl, T. K. (2010). Evaluation of the interactionist model of socioeconomic status and problem behavior: a developmental cascade across generations. *Dev Psychopathol, 22*(3), 695-713. doi:10.1017/S0954579410000374

Mersky, J. P., Janczewski, C. E., & Topitzes, J. (2017). Rethinking the measurement of adversity: Moving toward second-generation research on adverse childhood experiences. *Child maltreatment*, *22*(1), 58-68. doi:10.1177/1077559516679513

Milan, S., & Wortel, S. (2015). Family obligation values as a protective and vulnerability factor among low-income adolescent girls. *J Youth Adolesc, 44*(6), 1183-1193. doi:10.1007/s10964-014-0206-8

Moore, T. M., Kaczkurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., ... & Lahey, B. B. (2020). Criterion validity and relationships between alternative hierarchical dimensional models of general and specific psychopathology. *Journal of abnormal psychology*, *129*(7), 677.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (8th ed.)*. Los Angeles, California: Muthén & Muthén.

Noelke, C., McArdle, N., Baek, M., Huntington, N., Huber, R., Hardy, E., & Acevedo-Garcia, D. (2020). *Child Opportunity Index 2.0 Technical Documentation*. Retrieved from diversitydatakids.org/researchlibrary/research-brief/how-we-built-it

Orthner, D. K., Jones-Sanpei, H., & Williamson, S. (2004). The Resilience and Strengths of Low-Income Families. *Family Relations, 53*(2), 159-167. doi:10.1111/j.0022-2445.2004.00006.x

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *Br J Psychiatry*, *207*(1), 15-22. doi:10.1192/bjp.bp.114.149591

Pedersen, M. L., Jozefiak, T., Sund, A. M., Holen, S., Neumer, S. P., Martinsen, K. D., . . . Lydersen, S. (2021). Psychometric properties of the Brief Problem Monitor (BPM) in children with internalizing symptoms: examining baseline data from a national randomized controlled intervention study. *BMC Psychol, 9*(1), 185. doi:10.1186/s40359-021-00689-1

Peverill, M., Dirks, M. A., Narvaja, T., Herts, K. L., Comer, J. S., & McLaughlin, K. A. (2021). Socioeconomic status and child psychopathology in the United States: A meta-analysis of population-based studies. *Clin Psychol Rev, 83*, 101933. doi:10.1016/j.cpr.2020.101933

Piper, B. J., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of Brief Problem Monitor, an abbreviated form of the Child Behavior Checklist. *Psychiatry and Clinical Neurosciences, 68*(10), 759-767. doi:10.1111/pcn.12188

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychol Methods, 16*(2), 93-115. doi:10.1037/a0022658

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behav Res, 42*(1), 185-227. doi:10.1080/00273170701341316

Quon, E. C., & McGrath, J. J. (2014). Subjective socioeconomic status and adolescent health: A meta-analysis. *Health Psychology, 33*(5), 433-447. doi:10.1037/a0033716

Reid, A., Halgunseth, L. C., Espinosa-Hernandez, G., & Vasilenko, S. A. (2018). Sociocultural Influences on the Association Between Negative Romantic Experiences and Psychological Maladjustment in Mexican Adolescents. *Journal of Research on Adolescence, 28*(4), 888-901. doi:10.1111/jora.12375

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine, 90*, 24-31. doi:10.1016/j.socscimed.2013.04.026

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol Methods*, *21*(2), 137-150. doi:10.1037/met0000045

Rucker, D. D., & Galinsky, A. D. (2017). Social power and social class: conceptualization, consequences, and current challenges. *Current Opinion in Psychology, 18*, 26-30. doi:10.1016/j.copsyc.2017.07.028

Rucker, D. D., Galinsky, A. D., & Dubois, D. (2012). Power and consumer behavior: how power shapes who and what consumers value. *J Consum Psychol*, *22*, 352-368.

Sanchez, A. L., Jent, J., Aggarwal, N. K., Chavira, D., Coxe, S., Garcia, D., . . . Comer, J. S. (2022). Person-Centered Cultural Assessment Can Improve Child

Mental Health Service Engagement and Outcomes. *J Clin Child Adolesc Psychol, 51*(1), 1-22. doi:10.1080/15374416.2021.1981340

Schaefer, J. D., Moffitt, T. E., Arseneault, L., Danese, A., Fisher, H. L., Houts, R., . . . Caspi, A. (2018). Adolescent Victimization and Early-Adult Psychopathology: Approaching Causal Inference Using a Longitudinal Twin Study to Rule Out Noncausal Explanations. *Clin Psychol Sci*, *6*(3), 352-371. doi:10.1177/2167702617741381

Schwartz, S. J. (2007). The applicability of familism to diverse ethnic groups: a preliminary study. *J Soc Psychol, 147*(2), 101-118. doi:10.3200/SOCP.147.2.101-118

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychol Assess*, *31*(12), 1428-1441. doi:10.1037/pas0000623

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*(9), 1013.

Singh, G. K. (2003). Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health, 93*(7), 1137-1143. doi:10.2105/ajph.93.7.1137

Snyder, H. R., Young, J. F., & Hankin, B. L. (2019). Chronic Stress Exposure and Generation Are Related to the P-Factor and Externalizing Specific Psychopathology in Youth. *J Clin Child Adolesc Psychol, 48*(2), 306-315. doi:10.1080/15374416.2017.1321002

Tiet, Q. Q., Bird, H. R., Davies, M., Hoven, C., Cohen, P., Jensen, P. S., & Goodman, S. (1998). Adverse life events and resilience. *J Am Acad Child Adolesc Psychiatry, 37*(11), 1191-1200. doi:10.1097/00004583-199811000-00020

Tracy, M., Zimmerman, F. J., Galea, S., McCauley, E., & Stoep, A. V. (2008). What explains the relation between family poverty and childhood depressive

symptoms? *J Psychiatr Res, 42*(14), 1163-1175. doi:10.1016/j.jpsychires.2008.01.011

Tuomela, R., & Tuomela, M. (2005). Cooperation and trust in group context. *Mind & Society, 4*(1), 49-84. doi:10.1007/s11299-005-0008-5

Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2018). Effect of Foster Care Intervention on Trajectories of General and Specific Psychopathology Among Children With Histories of Institutional Rearing: A Randomized Clinical Trial. *JAMA Psychiatry*, *75*(11), 1137-1145. doi:10.1001/jamapsychiatry.2018.2556

Wade, M., Plamondon, A., & Jenkins, J. M. (2021). A Family Socialization Model of Transdiagnostic Risk for Psychopathology in Preschool Children. *Res Child Adolesc Psychopathol*, *49*(8), 975-988. doi:10.1007/s10802-021-00789-x

Waszczuk, M. A. (2021). The utility of hierarchical models of psychopathology in genetics and biomarker research. *World Psychiatry, 20*(1), 65-66. doi:10.1002/wps.20811

Watts, A. L., Makol, B. A., Palumbo, I. M., De Los Reyes, A., Olino, T. M., Latzman, R. D., . . . Sher, K. J. (2021). How Robust Is the p Factor? Using Multitrait-Multimethod Modeling to Inform the Meaning of General Factors of Youth Psychopathology. *Clinical Psychological Science*, 216770262110551. doi:10.1177/21677026211055170

Ziebold, C., Evans-Lacko, S., Andrade, M. C. R., Hoffmann, M., Fonseca, L., Barbosa, M., . . . Gadelha, A. (2021). Childhood poverty and mental health disorders in early adulthood: evidence from a Brazilian cohort study. *Eur Child Adolesc Psychiatry*. doi:10.1007/s00787-021-01923-2

Part 3. Critical Appraisal

A Clinical Psychoanalytic Evaluation of Transdiagnostic Approaches to Psychopathology

3.1 Abstract

In this chapter, I offer a clinically informed critique of the quantitative methods used in this thesis, and the field of quantitative psychiatry more broadly. The reflections are a result of the conflict between my clinical training and my research training prior to that, where I optimistically applied quantitative models to examine questions around measurement in mental health (see Constantinou et al., 2019; Constantinou et al., 2020). My experience of psychological measurement during my clinical training has alerted me to the messiness of the clinical situation and has forced me to evaluate some of my initial hopes for quantitative nosologies. Much of my thinking has been influenced by psychoanalytic readings and my more recent experience in psychoanalytic psychotherapy. My intention is not to pit quantitative approaches against psychoanalysis, research against practice, science against art. Such categories are false, and debates futile. My hope is to try and reconcile, or at least synthesise, the often-polarized views of my researcher and clinician identities. I evaluate quantitative nosologies using a clinical lens on three themes: methodology (e.g., limitations in the tools we use to estimate transdiagnostic factors), epistemology (e.g., biases in the way we conceptualize transdiagnostic factors), and application (e.g., limitations in applying transdiagnostic factors to practice).

3.2 Quantitative Approaches to Psychopathology

Quantitative approaches aim to classify how mental health problems or personality more broadly are structured (Blashfield, 2012). It is not enough to theorize that there are three types of personalities or difficulties (e.g., psychotic, borderline, and neurotic; Steiner, 2003). We need to test how these characteristics

arise in nature (Meehl, 1992), the same way a chemist would determine the type of element by testing its chemical makeup. While the chemist uses spectroscopy to infer the chemical bonds within matter based on how it behaves, the psychometrician uses factor analysis to infer the traits within a group of people based on how they respond to self-report measures. People respond in similar ways to certain groups of problems, e.g., people who meet the criteria for depressive disorders also tend to meet the criteria for anxiety disorders (Kessler, Chiu, Demler, & Walters, 2005). We can use people's responses to clusters of problems to infer their position on an underlying continuous trait that is assumed to tie these problems together (e.g., internalizing). From this, we can determine a taxonomy or 'periodic table' of traits that organize the types of mental health problems people can have.

Quantitative nosologists have applied two types of structures or 'period tables' to psychopathology data over the last two decades. The first is the higher-order model, in which there are at least two broader dimensions that tie together depressive, anxiety, somatic symptoms (e.g., internalizing) and antisocial behaviour and drug and alcohol dependence (e.g., externalizing; Krueger, Caspi, Moffit, & Silva, 1998). More recently, a higher-order 'general psychopathology' dimension was added to explain the fact that internalizing and externalizing dimensions also positively co-occur (e.g., people who score report more internalizing problems also tend to report more externalizing problems; Forbes et al., 2017).

The second type of structure applied to psychopathology is the bifactor model, which, like the higher-order model, includes general psychopathology, internalizing, and externalizing factors (Caspi et al., 2014; Lahey et al., 2012). However, the difference is that the general psychopathology factor is assumed to

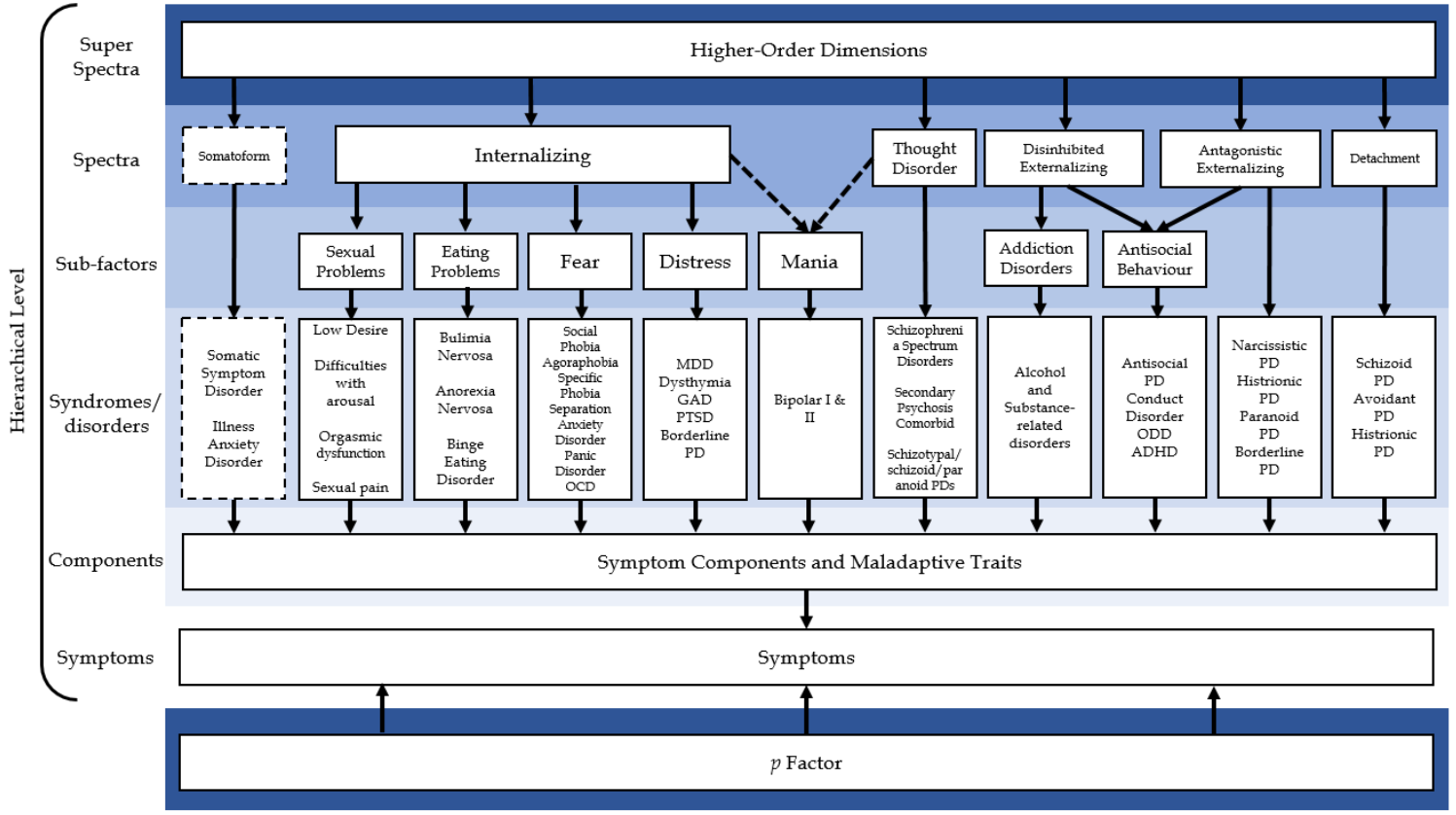
exist heterarchically alongside the internalizing and externalizing factors rather than hierarchically above them (Markon, 2019). That is, the general factor in the bifactor model reflects the way all symptoms tend to positively occur; the general factor in the higher-order model reflects the may internalizing and externalizing factors co-occur. Consequently, the internalizing and externalizing factors in the bifactor model reflect the ways in which certain groups of symptoms tend to co-occur, after taking into account what all symptoms have in common. The two models are statistically similar but have different theoretical implications (van Bork, Epskamp, Rhemtulla, Borsboom, & van der Maas, 2017) and methodological challenges (Forbes et al., 2021).

The higher-order model has inspired the 'Hierarchical Taxonomy of Psychopathology' (HiTOP), a conceptual map of the way in which symptoms coalescence into disorders, which in turn coalesce into narrower spectra up to the highest level of general psychopathology (Kotov et al., 2017). By contrast, the bifactor model has inspired a focus on the '*p*' factor, a single dimension that capture people's vulnerability to, or severity of, any and all forms of psychopathology (Caspi & Moffit, 2018). Both models are conceptually displayed in Figure 3.1. Theoretical and methodological differences aside, these models attempt to offer an alternative to current diagnostic nosologies which struggle to account for the high comorbidity rates among psychiatric diagnoses; high levels of heterogeneity within psychiatric diagnoses; the continuous rather than discrete nature of psychiatric difficulties; and the unreliability of clinical decisions regarding what constitutes a particular diagnosis and the threshold for meeting it (Kotov et al., 2021). Nonetheless, there has yet been a clinically informed response to quantitative nosologists about the challenge of applying quantitative models to clinical research

and practice. I will now outline issues in three domains–methodological, epistemological, and practical–and challenges associated with each from a psychodynamic lens.

Figure 3.1

Schematic of the higher-order (HiTOP) model and bifactor model adapted from Kotov et al. (2017)



3.3 Issue 1. Methodology

Transdiagnostic dimensions of psychopathology are currently estimated using self-report data from standardized questionnaires or interviews. The benefit of standardized instruments is that people's responses can be systematically catalogued and efficiently compared with each other (Cronbach, 1970). Nonetheless, standardized self-report measures raise several challenges for dynamic assessors and psychometricians alike. For instance, people do not just respond to scale items with their experience; responses are also affected their experience of the question, the response format, and measurement setting (Frost et al., 2007).

Self-report data are also affected by response styles. For example, people may underestimate the severity of their symptoms (compared to their observed level of functioning) due to a desire to present themselves in a socially desirable manner (Podsakoff, MacKenzie, & Podsakoff, 2012). Or they may over-estimate the severity of their symptoms due to a tendency to portray themselves negatively or to achieve some form of compensation (Podsakoff et al., 2012). People may also rely on the mid-point option (e.g., 'Not Sure'), perhaps because they feel ambivalently about the question, they lack insight (and are aware of this), they have trouble describing their internal states (i.e. alexithymia), or may be unmotivated to complete the assessment (Podsakoff et al., 2012).

Where the psychometrician and dynamic assessor differ is in what they do with these response styles. For the psychometrician, it is standard practice to view response styles as sources of noise that threaten the validity of item responses (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Like other forms of confounding, the goal is to assess response styles–say though symptom validity checks–and

statistically control for their impact on the 'true' score (Giromini, Young, & Sellbom, 2022). In contrast, for the dynamic assessor, the way someone responds gives meaning to the content of their responses; like the form and content of a poem, one cannot be understood without the other (Schafer, 1958). Response styles are shaped by the test situation which involves the respondent, tester, and their relationship. Both respondent and tester will bring phantasised and real reactions to the intimacy of assessing and being assessed, and their characteristic defences to manage the resulting anxieties (Schafer, 1954). Analyzing this rich source of information can deepen our understanding of test responses and hence the respondent's internal world.

The test situation cannot be underestimated. Experimental and social psychologists have paid dearly for overlooking the test situation and are struggling to replicate findings once taken as fact (Nosek et al., 2021). Some may ask: if the state-dependent aspects of the test situation are so influential, why are we able to estimate a *p* factor consistently over different test situations (see autoregressive coefficients in Part 2; McElroy, Belsky, Carragher, Fearon, & Patalay, 2018)? However, it is inaccurate to assume that the characteristics of the test situation are sporadic, as response styles also show consistency over time (Weijters, Geuens, & Schillewaert, 2010). What a psychometrician may treat as noise, a dynamic assessor treats as extensions of the respondent's ego organization, which is stable in nature. Furthermore, there is no reason why we cannot apply principles of reliability when assessing people's transient responses the test situation, e.g., by using a battery of tests and repeating tests over time (Schafer, 1953).

Using standardized self-report measures to estimate transdiagnostic factors poses a conundrum to the dynamic assessor. We are using *manifest* content as

indicators (e.g., people's conscious ratings) to estimate something like the *p* factor, which, ironically, is thought to reflect people's *latent* vulnerability to mental ill-health. Consider a patient who has a history of inpatient hospitalizations, and therefore shows a heightened vulnerability to mental health problems, but is functioning relatively well in the community. At the time of testing, they may report a low frequency and severity of symptoms, and hence a low *p* factor score, but this does not reflect their susceptibility to mental health difficulties should life turn for the worst.

The psychometrician will argue that even though we use manifest content like conscious ratings to estimate latent variables, they ultimately reflect broader constructs that are irreducible to their indicators (Harman, 1960). Nonetheless, the strength of this argument is challenged by findings that problem areas differ in how strongly they represent the *p* factor; a truly detached latent variable should be invariant to its indicators (Spearman, 1927). Moreover, the strength with which different problems reflect the *p* factor changes depending on the sample characteristics and measures used (see Chapter 1; Levin-Aspenson, Watson, Clark, & Zimmerman, 2021).

The relevance of response styles has in fact been acknowledged by quantitative nosologists. In their seminal paper, Lahey et al. (2012) suggested that even if the *p* factor reflects individual differences in a tendency to rate oneself unfavourably, this tendency may pose a general risk for common mental health problems (see also Seligman, Abramson, Semmel, & von Baeyer's [1979] depressive attributional style). Unfortunately, there have been few attempts to assess this hypothesis, despite it being an opportunity for mixed-methods research to assess both the content and form of responses in psychiatric interviews (see Fonagy,

Target, Steele, & Steele [1998] and Fonagy et al., [2016] for examples of how observational ratings and criterion keying, respectively, can be used to assess reflective functioning). Until more process-based measures are used to complement factor analytic models, our findings will be limited to the shortcomings of structured test responses.

3.4 Issue 2. Epistemology

Psychoanalysis and quantitative approaches to mental health both seek to understand how the ego/personality is organized across levels of health or functioning. However, they come to these understandings using methods linked to different epistemological traditions. Psychoanalysis has long favoured case studies of patients, literature, and more recently, institutions, to explore subjective meanings and idiosyncratic narratives, in line with post-modernist and idiographic traditions (Siegel, 2006). By contrast, quantitative approaches tend to use statistical methods to falsify hypotheses, infer causal relationships, and generalize findings to the wider population, in line with positivist and nomothetic traditions (Molenaar, 2004).

Clinicians may feel that quantitative approaches like HiTOP are too far removed from the clinical situation, lacking the richness and uniqueness of people's life stories. They may even berate quantitative approaches for lacking the essence of the unconscious and therefore have no place in psychoanalysis (Greene, 2000). I ask such readers to reflect on what might underpin their hesitancy or resentment, the same way they might reflect on unfavourable feelings towards a patient. After all, it is naïve to consider these approaches as diametrically opposed: quantitative research involves meaning-making in as much as case studies involve applying and

generalizing psychoanalytic theory (Luyten et al., 2006). Like Bornstein (2007), I consider these approaches to be different but complementary levels of analysis; the same way a concert musician uses both the conductor's feedback and their experience of the musicians around them to guide their playing.

There are, however, tensions between these levels of analysis that need reconciling. Psychiatric symptoms and disorders tend to positively co-occur at the nomothetic or between-person level. For instance, people who report high (or low) levels of internalizing symptoms also tend to report high (or low) levels of externalizing symptoms (Krueger et al., 1998). However, at the idiographic or within-person level, internalizing and externalizing symptoms can negatively co-occur or show non-linear associations (Reise, Ventura, Nuechterlein, & Kim, 2005). For example, someone who feels low and unmotivated may stabilize their mood through illicit substances, or they may use substances both when feeling low and when feeling euphoric. Positive associations among symptoms at the between-person level do not appear to capture the dynamic interplay among symptoms at the within-person level.

Negative or non-linear associations among symptoms at the idiographic level do not contradict positive associations at the nomothetic level, as the latter provides a coarse summary of the former. For example, someone who cycles between low mood and substance misuse throughout the month will report a high occurrence of both symptoms over the entire month. However, a *p* factor estimated at the between-person level cannot explain *why* people differ in their rates of comorbidity; that is fundamentally a within-person explanation (Curran & Bauer, 2010). Therefore, debates around whether *p* reflects a cause (i.e. latent vulnerability hypothesis; Caspi & Moffit, 2018) or a consequence (i.e. severity hypothesis; Smith,

Atkinson, Davis, Riley, & Oltmanns, 2020) of mental ill-health cannot be fully addressed at the between-person level.

A similar critique can be made of classical drive theory. For example, self-destructive behaviours inside the consulting room (e.g., the patient who gets worse after a seemingly enlightening interpretation) and outside of it (e.g., the patient who continuously enters relationships with abusive partners) may be explained by a death drive (Freud, 1920/1955). However, a death drive tells us little about why our patient appears to be drawn to abusive partners and not a harmful use of alcohol (though it might explain why another patient reports both). As a universal, between-person concept, a death drive explains everything yet nothing at all. We are also left needing to explain the origins of a death drive, Freud's original explanation for which is far removed from current neurobiological thought (Kernberg, 2009).

To investigate how and why someone develops the pattern of behaving and relating that they do, we require idiographic methods that map out the way that symptoms co-evolve for a person over time. Several statistical methods are available for analyzing covariation patterns in intensively sampled, personalized measures collected over time for single individuals, such as Vector Autoregression and Group Iterative Multiple Model Estimation (Beltz, Wright, Sprague, & Molenaar, 2016), Unified and Dynamic Factor Analysis (Piccirillo & Rodebaugh, 2019), and Personalized Network Models (Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017). These methods offer an opportunity to study the interplay of multiple symptoms at the individual level, as well as consistencies and differences at the group level, since idiographic and nomothetic are not distinct but are two complementary levels of analysis (Luyten et al., 2006; Wright & Zimmerman, 2019).

To put it analytically, idiographic methods offer transdiagnostic research what Melanie Klein offered drive theory. Klein (1957) suggested, among other things, that aggression directed towards the self and others may be a consequence of the phantasized frustrations that arise from the inevitable lapses or failures in caregiving. In this explanation, the construct of interest (e.g., destructiveness) is explained from the (perceived) interactions between variables (e.g., the caregiver and infant), like how idiographic approaches explain the constituents of transdiagnostic factors from the interactions between symptoms. There is no need to explain (away) hostility with biology, such as a death drive or underlying vulnerability.

Another tension linked to the positivist approach to mental health is the tendency to reify constructs. The *p* factor is no more than a summary variable of the associations among psychiatric problems (Constantinou & Fonagy, 2020), which does not necessarily reflect the causal processes underpinning these associations (Aristodemou & Fried, 2020). Yet, in our quest to find the neurobiological correlates of *p* (Zald & Lahey, 2017), there is a danger of treating transdiagnostic factors as if they exist in nature (Bonifay, Lane, & Reise, 2016). The same perversion occurs when we attempt to fit people into diagnostic categories or compare different brands of psychological therapy. In both cases, uncritical acceptance of our constructs blinds us the broader processes at play (e.g., shared mechanisms of distress or therapeutic change). Ironically, this is the kind of thinking that the *p* factor challenges because it represents processes that go beyond any single diagnostic entity. But the researcher is no less immune to the need for certainty than the clinician who labels their patients' predicament, and in doing so, distances themselves from the underlying complexity.

In placing too much faith in constructs like the *p* factor, we risk inadvertently locating the severity of problems within people.[7] We look for mechanisms within the individual, such as disordered thought processing (Caspi & Moffit, 2018) or impulsive responding to emotions (Carver, Johnson, & Timpano, 2017), and ignore contextual factors that influence people's hardship. Yet, higher *p* scores are associated with environmental stress (see Chapter 2; Lynch, Sunderland, Newton, & Chapman, 2021), parental mental health difficulties (Martel et al., 2017), racial and peer victimization (Liu, Mustanski, Dick, Bolland, & Kertes, 2017; Schaefer et al., 2018), and socio-economic disadvantage (see Part 2; Wade, Plamondon, & Jenkins, 2021).

The findings reviewed above suggest that the *p* factor partly captures variation in the lack of safety and stability in people's environments, rather than purely a vulnerability inherent within people (Luyten & Fonagy, 2022). This is not to deny the importance of the constitutional factors: children's self-regulatory and verbal abilities can moderate the association between socio-economic disadvantage and emotional and behavioural problems (Flouri, Midouhas, & Joshi, 2014).[8] However, we require a more transactional way of thinking about how the individual learns from and interacts with their environment (Fonagy et al., 2021; Sameroff, 2009). Put simply, there is no such thing as a *p* factor without a social context.

[7]This is partly a reflection of a neo-liberal tradition to quantify resources or assets in people, products, and markets.

[8]Of course, constructs such as self-regulation, verbal ability and the unconscious have undergone development in a particular environment, so they too are products of this person-context interaction.

3.5 Issue 3. Clinical Application

Quantitative approaches to mental health are poised to revolutionise practice by offering clinicians a more holistic way of classifying patients' difficulties (Kotov et al., 2021). Greater diagnostic precision could improve prognosis and encourage more tailored care to suit the needs of individual patients (Ruggero al., 2019). Hopwood et al. (2019) and Rodriguez-Seijas, Eaton, and Krueger (2015) outlined a stepped approach to applying HiTOP in practice, where the highest levels of the hierarchy (e.g., *p*) are assessed first. For instance, patients can be screened with a measure of general functioning, which can guide the intensity of treatment offered (e.g., from guided self-help to inpatient stay). Depending on time and resources, clinicians could work their way down the hierarchy to further customize the intervention. For example, spectral level measures such as internalizing and antagonistic/disinhibited externalizing could inform us about the potential nature of the therapeutic relationship and stance clinicians may need to take to ensure engagement (e.g., directive vs. non-directive). Furthermore, syndrome-based measures such as depression or sleep disturbances could inform the specific targets of treatment.

As it currently stands, the mechanisms that characterise higher-order dimensions like *p*, internalizing, and externalizing are not understood. Therefore, when we assess internalizing and externalizing in a patient with diagnoses of borderline personality disorder, generalized anxiety disorder, and depression (e.g., Rodriguez-Seijas et al., 2015), it is uncertain what we are in fact measuring. We might explain high levels of internalizing and externalizing as difficulties in emotion regulation and disinhibition which link these disorders together, but this is nothing more than re-expression of the symptomatology. Whilst there is an

advantage in assessing multiple problem domains, without a theoretical framework to understand the HiTOP dimensions, we are no better at describing the aetiology or mechanisms underlying a patient's profile of distress than we are when using traditional diagnoses. HiTOP prides itself on being theoretically agnostic (Kotov et al., 2017), perhaps to its detriment.

Some have argued that the Five Factor Model of Personality offers a theoretical framework for understanding the HiTOP dimensions (Widiger et al., 2018). Higher-order psychopathology dimensions may be analogous to the five factor traits (e.g., internalizing-neuroticism, detachment-extraversion, antagonism-agreeableness, disinhibition-conscientiousness, and openness to experience-psychoticism (Kotov et al., 2021). The profile of personality traits reflected in psychopathology dimensions can help us understand the 'character' of a patient's presenting difficulties (Shedler & Westen, 2007).

Take, for example, a patient who struggles to maintain romantic relationships. They may, among other things, report feeling worthless**,** easily upset, and that they are generally to blame, which indicates an internalizing/neurotic organization. Another patient with the same difficulty in maintaining romantic relationships might instead report being argumentative, critical of others, and at times physically aggressive, which indicates a more antagonistic externalizing/non-agreeable organization. This simplified example[9] highlights how two people may manage that the same problem in different ways and probably require different therapeutic interventions. However, both presentations may be underpinned by

[9]In practice, all trait domains would be evaluated.

common anxieties (e.g., fear of separation) represented at the level of *p*, which would also explain the way the same patient can oscillate between different ways of managing common anxieties (i.e. *p*) whilst defaulting to one defence more often (i.e. specific psychopathology factors).

Whilst profiling people on the HiTOP dimensions can give us a better sense of their difficulties, piecing together how people's scores relate their unique histories and relationships to ourselves and others requires the skill of clinical intuition (Shedler, 2015). Quantitative approaches to mental health pride themselves on being data-led and free from biases in judgement, but it is exactly this judgement and clinical sensibility that skilled clinicians use to understand people's tragedy (Schafer, 1967). Clinical intuition is often considered to be unreliable compared to statistical prediction, but the methods used to evaluate clinical judgement may themselves be unreliable; in more applied contexts, clinical judgement can be reliable and valid (Westen & Weinberger, 2004). There are methods that quantify clinical judgement, making it accessible to statistical analysis (e.g., Q-analysis, Shedler-Westen Assessment of Personality; Shedler & Westen, 1998), but HiTOP is yet to integrate them. For HiTOP and similar quantitative approaches to be integrated into practice, clinicians must feel that these approaches complement their skillset, rather than usurp it, as was implied by past clinical manuals (Shedler, 2015).

Another benefit of a hierarchical assessment of psychopathology is that we can examine the level(s) at which our intervention is taking effect. Current 'diagnosis-specific' interventions may in fact work by influencing transdiagnostic factors (Rodriguez-Seijas et al., 2015). For example, Selective Serotonin Reuptake Inhibitors (SSRIs), which were initially developed for depression, can be effective in treating anxiety disorders, eating disorders, personality disorders, and

schizophrenia (Vaswani, Linda, & Ramesh, 2003), suggesting that SSRIs impact broader dimensions like neuroticism/internalizing (Zemestani, Ommati, Rezaei, & Gallagher, 2022). Moreover, equivalence among psychotherapies (i.e. the Dodo bird verdict; Luborsky et al., 2002) may be a result of common therapeutic processes affecting the *p* factor (Fonagy, Luyten, Campbell, & Allison, 2014). If treatments mainly target broader psychopathology dimensions, then we may have more success in developing transdiagnostic interventions rather than disorder-specific interventions. However, Hopwood et al. (2019) are careful to point out that mapping specific interventions onto the HiTOP dimensions is somewhat misleading, as interventions are not applied purely in practice and are often blended or used interchangeably with aspects of other interventions.

An example of a transdiagnostic intervention is the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (Barlow et al., 2010) which targets shared processes across the internalizing spectrum, such as negative affectivity and emotional avoidance, with some success (Carlucci, Saggino, & Balsamo, 2021; Sakiris & Berle, 2019). I would argue, however, that there is no such thing as a transdiagnostic intervention: the distinction between "transdiagnostic" and "disorder-specific" is a matter of language rather than praxis. An intervention that targets common processes across disorders (e.g., emotion regulation) is still delivered through a specific means, often pre-existing techniques. For instance, intervention modules in the Unified Protocol are focused on emotions rather than symptoms but feature the core components of a disorder-specific CBT intervention, like psychoeducation on the links between thoughts, feelings, behaviours and the body; identifying maladaptive appraisals linked to physiological experiences; exposure to aversive experiences; and managing behavioural avoidance. What

differs is the framing of the difficulty (e.g., from symptoms to emotions) and the presentations it is applied to (e.g., beyond depression and anxiety to eating disorders, BPD, etc.). This would explain why the Unified Protocol shows equivalence, rather than superiority, to active controls like disorder-specific CBT in randomized controlled trials (Barlow et al., 2017; Eustis et al., 2020).

If quantitative approaches to psychopathology only offer a difference in language rather than praxis, we are left with a stark, perhaps cynical, conclusion: quantitative approaches do not fundamentally change practice. As Zimmerman (2021) put it, "I would expect that the treatment of most patients would be the same regardless of the diagnostic approach. For example, selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors will be prescribed whether patients were diagnosed with major depressive disorder or generalized anxiety disorder, or have elevated scores on an internalizing dimension with high scores on depression or fear subfactors" (p. 71). Along similar lines, we can ask whether practice has changed since routine outcome measures that capture transdiagnostic factors, like the Child Behavior Checklist (Achenbach & Rescorla, 2001) and Strengths and Difficulties Questionnaire (Goodman, 1997), became mainstream in child and adolescent mental health services (Callaghan, Fellin, & Warner-Gale, 2017; Hall et al., 2014). As Zimmerman noted, around half of people will get better regardless of how they are assessed, and a small but significant portion of service users will not respond to treatment.

To understand why some people do not respond to treatment, we need to go beyond current systems of thinking about diagnoses and therapies. All forms of assessment and intervention, including pharmacotherapies, occur within a relational context. Something is communicated between clinician and patient when

questions are asked in a clinical interview, an interpretation is offered in psychotherapy, or a pharmacologic agent is prescribed. The way the communicated message is experienced will be influenced by the person's earliest experiences of communication about themselves, the world, and the cultural sphere (Fonagy et al., 2021). For example, the patient who experienced a caregiver as overly involved and protective may experience a clinician who asks about the HiTOP spectra as prying and intrusive. Or the patient who was never allowed a say in the family may experience the calming effects of anxiolytics as silencing and restrictive.

If we want to change practice for the better and reach those who feel they have not been understood–or worse, may not even approach services–then understanding how people are oriented to communication is critical, as we are ultimately communicating a different understanding of their suffering and an alternative way of living (Allen, 2016; Fonagy & Allison, 2014). Some may develop an openness to learning from experience that challenges deep-seated phantasies. Others may continue to project their phantasies onto their experiences in a way that makes them mistrustful of, overly dependent on, or ambivalent to the messages communicated. The benefits of being understood by another are not reaped. A truly hierarchical analysis of communication would examine the way in which personal, familial, cultural, and socio-political forces shape a person's experience of communication.

3.6 Conclusions

In this essay, I have tried to critique modern approaches to quantifying psychopathology using more classical thought from the psychoanalytic tradition. The methodological, epistemological, and clinical issues discussed are all governed

by an overarching issue: the research-practice gap. At times I have presented this gap as insoluble, like two percepts of an illusion that cannot be seen simultaneously. At other times, I have presented methods and approaches that can bridge the gap, including process-based assessment, idiographic and transactional analyses, and a shift towards understanding the functional basis of our assessment and treatment methods. There is no reason why separatist and integrationist attitudes towards the research-practice gap should be opposing. It is not possible for research to fully capture the clinical context and vice versa, nor is it the goal I would argue. Both the randomized clinical trial and case-study serve different but equally useful purposes. The bifactor model not only offers us a way of operationalizing the general and specific characteristics of mental health, but it also offers us a way of thinking more dialectically, in that different levels of analysis can co-exist. The more we become aware of our current system of knowledge and practice, the more freedom we have to change it rather than repeat patterns of old.

3.7 References

Achenbach T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Allen, J. G. (2016). Should the century-old practice of psychotherapy defer to science and ignore its foundations in two millennia of ethical thought? *Bulletin of the Menninger Clinic*, *80*(1), 1.

Aristodemou, M. E., & Fried, E. I. (2020). Common Factors and Interpretation of the p Factor of Psychopathology. *J Am Acad Child Adolesc Psychiatry, 59*(4), 465-466. doi:10.1016/j.jaac.2019.07.953

Barlow, D. H., Farchione, T. J., Bullis, J. R., Gallagher, M. W., Murray-Latin, H., Sauer-Zavala, S., . . . Cassiello-Robbins, C. (2017). The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Compared With Diagnosis-Specific Protocols for Anxiety Disorders: A Randomized Clinical Trial. *JAMA Psychiatry, 74*(9), 875-884. doi:10.1001/jamapsychiatry.2017.2164

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & May, J. T. E. (2010). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. Oxford University Press.

Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. *Assessment, 23*(4), 447-458. doi:10.1177/1073191116648209

Blashfield, R. K. (2012). *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches*. Springer Science & Business Media.

Bornstein, R. F. (2007). Nomothetic psychoanalysis. *Psychoanalytic Psychology, 24*(4), 590-602. doi:10.1037/0736-9735.24.4.590

Callaghan, J. E., Fellin, L. C., & Warner-Gale, F. (2017). A critical analysis of Child and Adolescent Mental Health Services policy in England. *Clin Child Psychol Psychiatry, 22*(1), 109-127. doi:10.1177/1359104516640318

Carlucci, L., Saggino, A., & Balsamo, M. (2021). On the efficacy of the unified protocol for transdiagnostic treatment of emotional disorders: A systematic review and meta-analysis. *Clin Psychol Rev, 87*, 101999. doi:10.1016/j.cpr.2021.101999

Carver, C. S., Johnson, S. L., & Timpano, K. R. (2017). Toward a Functional View of the P Factor in Psychopathology. *Clin Psychol Sci, 5*(5), 880-889. doi:10.1177/2167702617710037

Caspi, A., & Moffitt, T. E. (2018). All for One and One for All: Mental Disorders in One Dimension. *Am J Psychiatry, 175*(9), 831-844. doi:10.1176/appi.ajp.2018.17121383

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*, *2*(2), 119-137. doi:10.1177/2167702613497473

Constantinou, M. P., & Fonagy, P. (2020). Drs. Constantinou and Fonagy Reply. *J Am Acad Child Adolesc Psychiatry, 59*(4), 466-467. doi:10.1016/j.jaac.2019.11.026

Constantinou, M., Frueh, B., Fowler, J., Allen, J., Madan, A., Oldham, J., & Fonagy, P. (2020). Predicting depression outcomes throughout inpatient treatment using the general and specific personality disorder factors. *Psychological Medicine*, 1-9. doi:10.1017/S003329172000361X

Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., … Fonagy, P. (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. doi:10.1016/j.jaac.2018.11.011

Cronbach, L. J. (1970). *Essentials of Psychological Testing*. New York: Harper & Row.

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

Eustis, E. H., Gallagher, M. W., Tirpak, J. W., Nauphal, M., Farchione, T. J., & Barlow, D. H. (2020). The Unified Protocol compared with diagnosis-specific protocols for anxiety disorders: 12-month follow-up from a randomized clinical trial. *Gen Hosp Psychiatry, 67*, 58-61. doi:10.1016/j.genhosppsych.2020.08.012

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *J Abnorm Psychol, 126*(8), 1044-1056. doi:10.1037/abn0000311

Flouri, E., Midouhas, E., & Joshi, H. (2014). The role of urban neighbourhood green space in children's emotional and behavioural resilience. *Journal of Environmental Psychology, 40*, 179-186. doi:10.1016/j.jenvp.2014.06.007

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, *51*(3), 372–380. doi:10.1037/a0036505

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2021). Culture and psychopathology: An attempt at reconsidering the role of social learning. *Dev Psychopathol*, 1-16. doi:10.1017/S0954579421000092

Fonagy, P., Luyten, P., Campbell, C., & Allison, L. (2014, December). *Epistemic trust, psychopathology and the great psychotherapy debate*. [Web Article]. Retrieved from http://www.societyforpsychotherapy.org/epistemic-trust-psychopathology-and-the-great-psychotherapy-debate

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., . . . Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLoS One, 11*(7), e0158678. doi:10.1371/journal.pone.0158678

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual, version 5.0, for application to adult attachment interviews*. London: University College London.

Forbes, M. K., Greene, A. L., Levin-Aspenson, H. F., Watts, A. L., Hallquist, M., Lahey, B. B., ... & Krueger, R. F. (2021). Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. *Journal of Abnormal Psychology*, *130*(3), 297-317.

Forbes, M. K., Kotov, R., Ruggero, C. J., Watson, D., Zimmerman, M., & Krueger, R. F. (2017). Delineating the joint hierarchical structure of clinical and personality disorders in an outpatient psychiatric sample. *Compr Psychiatry, 79*, 19-30. doi:10.1016/j.comppsych.2017.04.006

Freud, S. (1955). Beyond the pleasure principle. In *The standard edition of the complete psychological works of Sigmund Freud*, *Volume XVIII (1920-1922): Beyond the pleasure principle, group psychology and other works* (pp. 1-64).

Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & Mayo, F. D. A. P.-R. O. C. M. G. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health, 10 Suppl 2*, S94-S105. doi:10.1111/j.1524-4733.2007.00272.x

Giromini, L., Young, G., & Sellbom, M. (2022). Assessing Negative Response Bias Using Self-Report Measures: New Articles, New Issues. *Psychological Injury and Law, 15*(1), 1-21. doi:10.1007/s12207-022-09444-2

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, *38*(5), 581-586.

Green, A. (2002). Science and science fiction in infant research. In J. Sandler, A.-M. Sandler, & R. Davies (Eds.), *Clinical and Observational Psychoanalytic Research: Roots of a Controversy* (1st ed., pp. pp. 100–107). London: Karnac Books

Hall, C. L., Moldavsky, M., Taylor, J., Sayal, K., Marriott, M., Batty, M. J., . . . Hollis, C. (2014). Implementation of routine outcome measurement in child and adolescent mental health services in the United Kingdom: a critical perspective. *Eur Child Adolesc Psychiatry, 23*(4), 239-242. doi:10.1007/s00787-013-0454-2

Harman, H. H. (1960). *Modern Factor Analysis*. Chicago: University of Chicago Press

Hopwood, C. J., Bagby, R. M., Gralnick, T., Ro, E., Ruggero, C., Mullins-Sweatt, S., . . . Zimmermann, J. (2020). Integrating psychotherapy with the hierarchical taxonomy of psychopathology (HiTOP). *Journal of Psychotherapy Integration, 30*(4), 477-497. doi:10.1037/int0000156

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78*(3), 229-248. doi:10.1037/h0030852

Kernberg, O. (2009). The concept of the death drive: a clinical perspective. *Int J Psychoanal, 90*(5), 1009-1023. doi:10.1111/j.1745-8315.2009.00187.x

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 617-627. doi:10.1001/archpsyc.62.6.617

Klein, M. (1957). *Envy and gratitude*. London, UK: Hogarth Press.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol*, *126*(4), 454-477. doi:10.1037/abn0000258

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., . . . Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annu Rev Clin Psychol*, *17*, 83-108. doi:10.1146/annurev-clinpsy-081219-093304

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), 216-227. doi:10.1037/0021-843x.107.2.216

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? *J Abnorm Psychol*, *121*(4), 971-977. doi:10.1037/a0028355

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. doi:10.1177/1073191120954921

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychol Methods*, *22*(3), 486-506. doi:10.1037/met0000075

Luborsky, L. (2002). The Dodo Bird Verdict Is Alive and Well--Mostly. *Clinical Psychology: Science and Practice, 9*(1), 2-12. doi:10.1093/clipsy/9.1.2

Luyten, P., Blatt, S. J., & Corveleyn, J. (2006). Minding the gap between positivism and hermeneutics in psychoanalytic research. *J Am Psychoanal Assoc, 54*(2), 571-610. doi:10.1177/00030651060540021301

Luyten, P., & Fonagy, P. (2022). Integrating and differentiating personality and psychopathology: A psychodynamic perspective. *J Pers, 90*(1), 75-88. doi:10.1111/jopy.12656

Lynch, S. J., Sunderland, M., Newton, N. C., & Chapman, C. (2021). A systematic review of transdiagnostic risk and protective factors for general and specific psychopathology in young people. *Clin Psychol Rev, 87*, 102036. doi:10.1016/j.cpr.2021.102036

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosario, M. C., Mari, J. J., . . . Salum, G. A. (2017). A general psychopathology factor (P factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *J Abnorm Psychol, 126*(1), 137-148. doi:10.1037/abn0000205

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annu Rev Clin Psychol*, *15*, 51-69. doi:10.1146/annurev-clinpsy-050718-095522

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? *J Child Psychol Psychiatry, 59*(6), 667-675. doi:10.1111/jcpp.12849

Meehl, P. E. (1992). Factors and Taxa, Traits and Types, Differences of Degree and Differences in Kind. *Journal of Personality, 60*(1), 117-174. doi:10.1111/j.1467-6494.1992.tb00269.x

Miller, M. W., Fogler, J. M., Wolf, E. J., Kaloupek, D. G., & Keane, T. M. (2008). The internalizing and externalizing structure of psychiatric comorbidity in combat veterans. *J Trauma Stress, 21*(1), 58-65. doi:10.1002/jts.20303

Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, and standards. *Journal of Abnormal Psychology, 100*(3), 245-261. doi:10.1037/0021-843x.100.3.245

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 & Perspective, 2*(4), 201-218. doi:10.1207/s15366359mea0204_1

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., . . . Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annu Rev Psychol, 73*, 719-748. doi:10.1146/annurev-psych-020821-114157

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clin Psychol Rev, 71*, 90-100. doi:10.1016/j.cpr.2019.01.002

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol, 88*(5), 879-903. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol, 63*, 539-569. doi:10.1146/annurev-psych-120710-100452

Reise, S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *J Pers Assess, 84*(2), 126-136. doi:10.1207/s15327752jpa8402_02

Rodriguez-Seijas, C., Eaton, N. R., & Krueger, R. F. (2015). How Transdiagnostic Factors of Personality and Psychopathology Can Inform Clinical Assessment and Intervention. *J Pers Assess, 97*(5), 425-435. doi:10.1080/00223891.2015.1055752

Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., . . . Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *J Consult Clin Psychol, 87*(12), 1069-1084. doi:10.1037/ccp0000452

Sakiris, N., & Berle, D. (2019). A systematic review and meta-analysis of the Unified Protocol as a transdiagnostic emotion regulation based intervention. *Clin Psychol Rev, 72*, 101751. doi:10.1016/j.cpr.2019.101751

Sameroff, A. (2009). The transactional model. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 3–21). American Psychological Association. https://doi.org/10.1037/11877-001

Schaefer, J. D., Moffitt, T. E., Arseneault, L., Danese, A., Fisher, H. L., Houts, R., . . . Caspi, A. (2018). Adolescent Victimization and Early-Adult Psychopathology: Approaching Causal Inference Using a Longitudinal Twin Study to Rule Out Noncausal Explanations. *Clin Psychol Sci*, *6*(3), 352-371. doi:10.1177/2167702617741381

Schafer, R. (1953). Content analysis in the Rorschach test. *J Proj Tech, 17*(3), 335-339. doi:10.1080/08853126.1953.10380496

Schafer, R. (1954). Some applications of contemporary psychoanalytic theory to projective testing. *J Proj Tech, 18*(4), 441-447. doi:10.1080/08853126.1954.10380581

Schafer, R. (1958). How was this story told. *J Proj Tech, 22*(2), 181-210. doi:10.1080/08853126.1958.10380840

Schafer, R. (1967). *Projective testing and psychoanalysis*. New York, NY: International Universities Press.

Seligman, M. E., Abramson, L. Y., Semmel, A., & von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology*, *88*(3), 242–247. https://doi.org/10.1037/0021-843X.88.3.242

Shedler, J. (2015). Integrating clinical and empirical perspectives on personality: The Shedler-Westen Assessment Procedure (SWAP). 225-252. doi:10.1037/14549-010

Shedler, J., & Westen, D. (1998). Refining the measurement of axis II: a Q-sort procedure for assessing personality pathology. *Assessment, 5*(4), 333-353. doi:10.1177/107319119800500403

Shedler, J., & Westen, D. (2007). The Shedler-Westen Assessment Procedure (SWAP): making personality diagnosis clinically meaningful. *J Pers Assess, 89*(1), 41-55. doi:10.1080/00223890701357092

Siegel, E. (2005). Psychoanalysis as a traditional form of knowledge: an inquiry into the methods of psychoanalysis. *International Journal of Applied Psychoanalytic Studies, 2*(2), 146-163. doi:10.1002/aps.38

Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The General Factor of Psychopathology. *Annu Rev Clin Psychol*, *16*, 75-98. doi:10.1146/annurev-clinpsy-071119-115848

Spearman, C. (1927). *The abilities of man*. Oxford, England: Macmillan

Steiner, J. (2003). *Psychic retreats: Pathological organizations in psychotic, neurotic and borderline patients*. Routledge.

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, *27*(6), 759-773. doi:10.1177/0959354317737185

Vaswani, M., Linda, F. K., & Ramesh, S. (2003). Role of selective serotonin reuptake inhibitors in psychiatric disorders: a comprehensive review. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 27*(1), 85-102. doi:10.1016/s0278-5846(02)00338-x

Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2018). Effect of Foster Care Intervention on Trajectories of General and Specific Psychopathology Among Children With Histories of Institutional Rearing: A Randomized Clinical Trial. *JAMA Psychiatry*, *75*(11), 1137-1145. doi:10.1001/jamapsychiatry.2018.2556

Wade, M., Plamondon, A., & Jenkins, J. M. (2021). A Family Socialization Model of Transdiagnostic Risk for Psychopathology in Preschool Children. *Res Child Adolesc Psychopathol*, *49*(8), 975-988. doi:10.1007/s10802-021-00789-x

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychol Methods, 15*(1), 96-110. doi:10.1037/a0018721

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *Am Psychol, 59*(7), 595-613. doi:10.1037/0003-066X.59.7.595

Widiger, T. A., Sellbom, M., Chmielewski, M., Clark, L. A., DeYoung, C. G., Kotov, R., . . . Wright, A. G. C. (2018). Personality in a Hierarchical Model of Psychopathology. *Clinical Psychological Science, 7*(1), 77-92. doi:10.1177/2167702618797105

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychol Assess, 31*(12), 1467-1480. doi:10.1037/pas0000685

Zald, D. H., & Lahey, B. B. (2017). Implications of the Hierarchical Structure of Psychopathology for Psychiatric Neuroimaging. *Biol Psychiatry Cogn Neurosci Neuroimaging, 2*(4), 310-317. doi:10.1016/j.bpsc.2017.02.003

Zemestani, M., Ommati, P., Rezaei, F., & Gallagher, M. W. (2022). Changes in neuroticism-related constructs over the Unified Protocol for Transdiagnostic

Treatment of Emotional Disorders in patients on an optimal dose of SSRI. *Personality disorders*, *13*(2), 171–181. doi:10.1037/per0000482

Zimmerman M. (2021). Why hierarchical dimensional approaches to classification will fail to transform diagnosis in psychiatry. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *20*(1), 70–71. doi:10.1002/wps.20815

Appendix 1. Literature Review

1.1 Model-based Reliability Indices

1.1.1 Explained Common Variance (*ECV*)

ECV can be calculated from a standardized factor loading matrix made up of a general factor (*G*) and three specific factors (S_1, S_2, S_3) as follows:

$$ECV = \frac{(\sum \lambda_G^2)}{(\sum \lambda_G^2) + (\sum \lambda_{S_1}^2) + (\sum \lambda_{S_2}^2) + (\sum \lambda_{S_3}^2)},$$

where the sum of squared general factor loadings is divided by the sum of squared general and specific factor loadings (i.e. the total modelled variance; Rodriguez, Reise, & Haviland, 2016b).

ECV-subscale can be used to compute the proportion of modelled variance explained by a given specific factor, in this case S_1, relative to the variance explained by all factors, for example:

$$ECV_s = \frac{(\sum \lambda_{S_1}^2)}{(\sum \lambda_G^2) + (\sum \lambda_{S_1}^2) + (\sum \lambda_{S_2}^2) + (\sum \lambda_{S_3}^2)},$$

where the sum of squared factor loadings for a specific factor of interest is divided by the total modelled variance (Stucky & Edelen, 2015).

1.1.2 Coefficient Omega (ω)

Omega can be calculated from a standardized factor loading matrix made up of a general factor (*G*) and three specific factors (S_1, S_2, S_3) as follows:

$$\omega = \frac{(\sum \lambda_G)^2 + (\sum \lambda_{S_1})^2 + (\sum \lambda_{S_2})^2 + (\sum \lambda_{S_3})^2}{(\sum \lambda_G)^2 + (\sum \lambda_{S_1})^2 + (\sum \lambda_{S_2})^2 + (\sum \lambda_{S_3})^2 + (\sum 1 - h^2)},$$

where the squared sum of factor loadings for the general and specific factors (i.e. the common variance) is divided by the squared sum of the common variance plus the unique or error variance (i.e. total variance).

1.1.3 Omega hierarchical (ωH)

Suppose we wanted to determine the proportion of variance in observed total scores attributable to the general factor alone. Using the same example above with one general factor and three specific factors, we could calculate omega hierarchical as follows:

$$\omega_H = \frac{(\sum \lambda_G)^2}{(\sum \lambda_G)^2 + (\sum \lambda_{S_1})^2 + (\sum \lambda_{S_2})^2 + (\sum \lambda_{S_3})^2 + (\sum 1 - h^2)},$$

where the squared sum of factor loadings for the general factor alone is divided by the total variance, with the variance associated with specific factors now treated as error. We can also determine the proportion of reliable variance (i.e. error-free variance) in total scores attributable to the general factor by dividing omega hierarchical by omega, which is known as relative omega (Dueber, 2017; Rodriguez et al., 2016b).

These principles can also be applied to specific factors. For example, we can calculate the proportion of variance in observed subscale scores attributable to the general factor and a specific factor of interest with omega-subscale:

$$\omega_S = \frac{(\sum \lambda_G)^2 + (\sum \lambda_{S_i})^2}{(\sum \lambda_G)^2 + (\sum \lambda_{S_i})^2 + (\sum 1 - h^2)}.$$

Similarly, we can determine the proportion of variance in observed subscale scores attributable to a given specific factor while controlling for the general factor using omega hierarchical-subscale (*ωHS*):

$$\omega_{HS} = \frac{(\sum \lambda_{S_i})^2}{(\sum \lambda_G)^2 + (\sum \lambda_{S_i})^2 + (\sum 1 - h^2)}.$$

Again, by dividing omega hierarchical-subscale by omega-subscale, we can determine the relative omega-specific, i.e. the proportion of reliable variance in subscale scores attributable to a specific factor of interest.

1.1.4 Construct Reliability *(H)*

H can be estimated for a given factor using standardized factor loadings as follows:

$$H = 1/\left[1 + \frac{1}{\sum_{i=1}^{K} \frac{\lambda_i^2}{1 - \lambda_i^2}}\right],$$

where the denominator reflects the sum of the ratios between each item's squared factor loading (i.e. proportion of variance explained by the factor) and the complement of each item's squared factor loading (i.e. the proportion of variance unexplained explained by the factor). The other calculations ensure that *H* ranges from 0-1, with higher scores indicating that a factor is more well-defined by a given set of indicators. *H* increases as the number of items and their loading strength increases.

1.1.5 Factor Score Determinacy *(FD)*

Factor scores are inherently 'indeterminate', i.e. for any set of factor scores estimated, there exists a different set of scores that could be derived from the same factor loading matrix (Gutmann, 1955). Factor determinacy (FD) is the reliability of factor scores, or the extent to which factor scores are good estimates of individual differences on the factor (Grice, 2001). FD can be calculated from the model-implied correlation matrix with the following formula:

$$FD = diag(\mathbf{\Phi \Lambda}^T \mathbf{\Sigma}^{-1} \mathbf{\Lambda \Phi})^{1/2},$$

where Φ is a $k \times k$ matrix of factor correlations, Λ is a $j \times k$ factor loading matrix, and Σ is a $k \times \underline{k}$ matrix of the model-implied factor correlations. FD values range from 0-1 and represent the correlation between the factor and factor scores.

1.1.6 Percentage of Uncontaminated Correlations *(PUC)*

PUC is estimated by first calculating the number of unique correlations among indicators, $p(p-1)/2$, where p is the number of indicators. In the example with one general factor and three specific factors, if there were 12 items, then there will be 66 unique correlations ([12*11]/2). Correlations between items within a specific factor (i.e. 'within-factor' correlations) can be explained by both the general and specific factors. However, correlations between indicators from different specific factors (i.e. 'between-factor' correlations) can only be explained by the general factor. We can calculate the proportion of unique correlations that are solely explained by the general factor, relative to the correlations explained by the general and specific factors using the following formula:

$$PUC = 1 - \frac{\sum_{i=1}^{3} S_i(p_i(p_i - 1)/2)}{p(p-1)/2},$$

where the numerator reflects the number of unique correlations that can be explained by the specific (and general) factors, and the denominator reflects the total number of unique correlations. The remainder of correlations would naturally be informed by the general factor alone, so we take the complement of the proportion. Continuing our example, if there were three specific factors with four loadings each, then 18 of the 66 unique correlations would be explained by the specific factors. Therefore, 27% of the unique correlations are 'contaminated' by multidimensionality, leaving 73% of the unique correlations to be explained by the general factor alone.

PUC is largest when there are multiple specific factors, each with a small number of loadings, because this increases the number of between-factor correlations that can be uniquely explained by the general factor compared to the number of within-factor correlations can be explained by both the general and specific factors. See Appendix 1.1 for a more detailed description of the model-based indices.

1.2 Model Description

1.2.1 Meta-Analysis

Fixed-effect models assume that each study approximates one, true reliability parameter in the population (e.g., the *p* factor has a fixed level of reliability in the population). Studies vary in their reliability estimates because of sampling and measurement error; no study captures the population parameter perfectly. However, this within-study variability is minimized when estimates are

aggregated across multiple studies, which ultimately converge towards the population parameter. The observed reliability coefficient, $\hat{\rho}$, for a given study, i, in a fixed-effect model is as follows:

$$\hat{\rho}_i = \rho + \varepsilon_i,$$

where ρ is the true reliability score in the population and ε_i is the within-study error (i.e. the difference between the true and observed reliability scores due to sampling and measurement error).

By contrast, random-effects models assume that each study approximates a reliability parameter from a range of possible populations (e.g., the *p* factor's reliability can vary depending on the population it is sampled from). There is no single population reliability, but multiple reliabilities from a range of populations. There may be overlap between these populations, but the aim is to estimate a distribution of true reliability coefficients from a randomly selected sample of studies that vary in their estimates due to 1) varying population characteristics (i.e. between-study variability) and 2) sampling and measurement error (i.e. within-study variability). The observed reliability coefficient in a random-effects model is expressed as follows:

$$\hat{\rho}_i = \mu_\rho + \xi_i + \varepsilon_i,$$

where μ_ρ is the mean of the population reliability parameters and ξ_i is the between-study variability (i.e. the difference between the population estimate and the population mean). The between-study variability is assumed to be normally distributed with a variance of τ^2, e.g., $\xi_i \sim N(0, \tau^2)$. The within-study variability, ε_i, is

the same as in the fixed-effect model (i.e. the difference between the true and observed reliability scores for a given study).

Random-effects models are multilevel in nature as they incorporate multiple, nested sources of variability. Take, for instance, the random-effects model described above. There are two sources of variance: the within-study variance and between-study variance. We can think of these sources as two levels of analysis. At the lowest level, there is the variability within each study (level 1). Each study also serves as a larger unit that can be compared with other studies (level 2). Therefore, a standard random-effects meta-analysis is equivalent to a two-level multilevel model.

Some studies in this meta-analysis reported multiple factor loading matrices, either because they ran a bifactor analysis for multiple samples or the same sample at different time-points. Analyzing each study as independent violates our model assumptions and can bias standard errors due to dependences within studies. We can, however, model these dependencies as another level of analysis in our meta-analysis. For instance, when analysing the model-based reliabilities of *p* factors across studies, each study measures *p* with its own level of error (e.g., sampling and measurement error; level 1), some studies will contribute multiple *p* factor estimates which will vary among themselves (level 2), and all studies will vary in their *p* factor estimates due to sampling for multiple populations (level 3). We therefore have a three-level random-effects model, which can be expressed as follows:

Level 1: Within-study variability

$$\hat{\rho}_{ij} = \rho_{ij} + \varepsilon_{ij},$$

which is the same as the fixed-effect estimate described above, except that we use ρ_{ij} rather than ρ to denote that there are multiple 'true' effect sizes for each study, i, within each time-point/sample reported by a study, j. ε_{ij} reflects the level 1 variability within each study estimate due to sampling and measurement error.

Level 2: Between sample/time-point variability

$$\rho_{ij} = \kappa_j + \xi_{ij},$$

where κ_j is the reliability estimate pooled across time-points/samples for a given study, and ξ_{ij} reflects the level 2 variability among time-points/samples within a given study (i.e. within-study heterogeneity).

Level 3: Between-study variability

$$\kappa_j = \mu + \xi_j,$$

where μ is population reliability averaged across studies and ξ_j is the level 3 variability across clusters of time-points/samples (i.e. between-study heterogeneity).

Notice how each estimate of $\hat{\rho}_{ij}$ is defined at each additional level, e.g., ρ_{ij} is defined by $\kappa_j + \xi_{ij}$, and κ_j is defined by $\mu + \xi_j$. The fully nested structure of a multilevel (random-effects) model for a given reliability estimate for the *p* factor is expressed as:

$$\hat{\rho}_{ij} = \mu + \xi_j + \xi_{ij} + \varepsilon_{ij}.$$

Each study also reported factor loadings for at least two specific factors. This adds an additional source of variability (e.g., variability between specific factor

reliability estimates within a study). I therefore estimated a four-level model for specific factors, summarized as follows:

$$\hat{\rho}_{isj} = \mu + \xi_j + \xi_{sj} + \xi_{isj} + \varepsilon_{isj,}$$

where μ is the population reliability averaged over specific factors across studies, ξ_j is the level 4 variability across clusters of time-points/samples (hence between studies); ξ_{sj} is the level 3 variability among time-points/samples within a given study (i.e. between-sample heterogeneity); ξ_{isj} is the level 2 variability among specific factors within a time-point/sample for a given study (i.e. between-specific factor heterogeneity); and $\varepsilon_{isj,}$ is the variability within each study estimate due to sampling and measurement error (i.e. within-study heterogeneity, which is synonymous with the variability within a specific factor estimate).

To summarize, I estimated a three-level model for each *p* factor reliability estimates, with random effects around the study estimates and the sample/time-point and a four-level model for the specific factor reliability estimates, with random effects around the study estimates, sample/time-point estimates, and specific factors.

1.2.2 Meta-Regression

In meta-regression, we regress the observed effect size, in the case a reliability coefficient, ρ, onto study-level predictors (continuous or categorical). We can add predictors to the three-level random-effects model predicting the *p* factor reliabilities as follows:

$$\hat{\rho}_{ij} = \mu + \beta_{x1_j} \ldots + \beta_{xk_j} + \xi_j + \xi_{ij} + \varepsilon_{ij},$$

where μ is population reliability pooled across studies, $\beta_{x_{1j}}$ is the predicted change in *p* factor reliability with a one-unit increase in predictor $x1_j$ holding *k* predictors constant, for study, j, ξ_j is the between-study (level 3) variability, ξ_{ij} is the between-sample (level 2) variability, and ε_{ij} is the within-study (level 1) variability.

This is technically a mixed-effects multilevel meta-regression model, as it includes both fixed-effects (e.g., β_{x_j}, since predictors are 'fixed' for each study) and random-effects (e.g., ξ_j, ξ_{ij}). Predictors could also vary at the between-sample level (e.g., differences between samples or time-varying predictors within a given study or $\beta_{x_{ij}}$) though I did not include any in my models. By introducing fixed-effect predictors into the model, we aim to reduce or explain the random-effects, hence between-study heterogeneity (e.g., τ^2). We can do the same thing for the four-level random effects model predicting variability in specific factor reliabilities:

$$\hat{\rho}_{isj} = \mu + \beta_{x1_j} \ldots + \beta_{xk_j} + \xi_j + \xi_{sj} + \xi_{isj} + \varepsilon_{isj,}$$

where in addition to the coefficients described above, ξ_j is the between-study (level 4) variability, ξ_{sj} is the between-sample (level 3) variability, ξ_{isj} is the between-specific factor (level 2) variability, and $\varepsilon_{isj,}$is the within-study (level 1) variability.

Table A1.1

Pooled p factor reliability estimates (ρ), including Bonett transformed (BT), Hakstian-Whallen T transformed (T), and raw estimates

Reliability Index	ρ	95% CI	95% PI	Q	Level	I^2	τ^2
ECV							
BT	.61***	.58-.64	.25-.80	15,677.10***	1	.34%	0.04×10^{-2}
					2	21%	0.02
					3	79%	0.09
T	.61***	.58-.64	.29-.81	14,920.81***	1	.29%	0.05×10^{-3}
					2	21%	0.01×10^{-1}
					3	79%	0.05×10^{-1}
Raw	.59***	.56-.63	.34-.85	14,180.05***	1	.34%	0.05×10^{-4}
					2	23%	0.04×10^{-1}
					3	77%	0.01
ω							
BT	.94***	.92-.95	.62-.99	65,312.56***	1	.05%	0.04×10^{-2}
					2	3%	0.03
					3	97%	0.82
T	.93***	.91-.94	.72-1	82,243.46***	1	.04%	0.06×10^{-5}
					2	3%	0.05×10^{-2}
					3	97%	0.01
Raw	.91***	.89-.93	.76-1	135,609.17***	1	.01%	0.02×10^{-2}
					2	3%	0.06×10^{-1}
					3	97%	0.01
ω_H							
BT	.78***	.74-.81	.19.-94	44,070.81***	1	.09%	0.04×10^{-2}

					2	14%	0.37
					3	85%	0.06
	T	.77***	.73-.80	.36-.95	50,220.82***	1	.10%	0.05×10^{-3}
						2	16%	0.03×10^{-1}
						3	84%	0.01
	Raw	.74***	.70-.77	.46-1	84,197.13***	1	.05%	0.01×10^{-3}
						2	19%	0.04×10^{-1}
						3	81%	0.02
Relative ω								
	BT	0.84***	.81-.87	.37-.96	54,865.53***	1	.08%	0.03×10^{-2}
						2	19%	0.10
						3	80%	0.39
	T	0.83***	.80-.85	.51-.97	58,486.29***	1	.09%	0.01×10^{-3}
						2	23%	0.03×10^{-1}
						3	76%	0.01
	Raw	0.81***	.78-.83	.59-1	92,694.53***	1	.04%	0.05×10^{-4}
						2	31%	0.04×10^{-1}
						3	69%	0.01
H								
	BT	.92***	.90-.94	.51-.99	75,132.45***	1	.04	0.04×10^{-2}
						2	7%	0.06
						3	93%	0.83
	T	.91***	.89-.93	.66-.99	90,618.09***	1	.04%	0.07×10^{-3}
						2	6%	0.01×10^{-1}
						3	94%	0.02

Raw	.89***	.87-.91	.72-1	151,500.46***	1	.01%	0.09×10^{-4}
					2	3%	0.02×10^{-2}
					3	97%	0.07×10^{-1}
FD							
BT	.95***	.94-.96	.69-.99	103,399.53***	1	.04%	0.04×10^{-2}
					2	45%	0.38
					3	55%	0.47
T	.94***	.93-.95	.77-.99	89,660.33***	1	.05%	0.06×10^{-3}
					2	41%	0.05×10^{-1}
					3	59%	0.08×10^{-1}
Raw	.93***	.91-.94	.82-1	88,682.21***	1	.02%	0.05×10^{-3}
					2	34%	0.01×10^{-1}
					3	66%	0.02×10^{-1}

*N*ote. Reliability estimates, confidence intervals, and prediction intervals have been back-transformed to the original reliability index scale. Heterogeneity estimates (τ^2) remain in the transformed (or raw) scale. There were 97 entries at level 1/2, and 68 at level 3. CI = confidence interval; *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; I^2 = percentage of between-study heterogeneity; ω = Coefficient Omega; ω_H = Coefficient Omega Hierarchical; ρ = pooled reliability estimate; PI = prediction interval; *Q* = Cochran's *Q* statistic; $\boldsymbol{\tau^2}$ **=** between-study variance component.

***$p < .001$

**$p < .01$

*$p < .05$

Table A1.2

Pooled specific psychopathology factor reliability estimates (ρ), including Bonett transformed (BT), Hakstian-Whallen T transformed (T), and raw estimates

Reliability Index	ρ	95% CI	95% PI	Q	Level	I^2	τ^2
ECVs							
BT	.16***	.14-.17	0-.31	11,957.13***	1	3%	0.03×10^{-2}
					2	28%	0.03×10^{-1}
					3	0%	0.07
					4	69%	0.07×10^{-1}
T	.15***	.14-.17	0-.31	12,237.94***	1	3%	0.03×10^{-3}
					2	29%	0.03×10^{-2}
					3	0%	0.02×10^{-10}
					4	69%	0.07×10^{-2}
Raw	.15***	.14-.17	0-.32	12,723.25***	1	3%	0.02×10^{-2}
					2	28%	0.02×10^{-1}
					3	0%	0.07×10^{-10}
					4	68%	0.05×10^{-1}
ω_S							
BT	.87***	.85-.89	.48-.97	212,515.96***	1	.06%	0.02×10^{-2}
					2	36%	0.17
					3	0%	0.01×10^{-9}
					4	64%	0.31
T	.86***	.84-.88	.57-.98	226,561.35***	1	.05%	0.08×10^{-4}
					2	29%	0.04×10^{-1}
					3	0%	0.04×10^{-10}
					4	71%	0.01

Raw	.84***	.81-.86	.61-1	254,470.48***	1	.02%	0.03×10^{-4}
					2	18%	0.02×10^{-1}
					3	0%	0.03×10^{-10}
					4	82%	0.01
ω_{HS}							
BT	.39***	.34-.42	0-.71	212,373.25***	1	.19%	0.03×10^{-2}
					2	74%	0.11
					3	0%	0.07×10^{-9}
					4	26%	0.20
T	.37***	.33-.41	0-.71	234,204.44***	1	.22%	0.02×10^{-3}
					2	72%	0.07×10^{-1}
					3	0%	0.01×10^{-9}
					4	27%	0.03×10^{-1}
Raw	.35***	.32-.39	0-.74	270,881.10***	1	.21%	0.08×10^{-3}
					2	72%	0.03
					3	0%	0.01×10^{-9}
					4	28%	0.01
Relative ω_S							
BT	.49***	.43-.54	0-.87	424,080.07***	1	.06%	0.03×10^{-2}
					2	81%	0.40
					3	0%	0.06×10^{-8}
					4	19%	0.09
T	.46***	.41-.51	0-.85	592,669.03***,	1	.09%	0.02×10^{-3}
					2	78%	0.02
					3	0%	0.02×10^{-9}
					4	22%	0.05×10^{-1}
Raw	.42***	.38-.46	0-.88	1,045,600.15***	1	.02%	0.01×10^{-3}

					2	76%	0.04
					3	0%	0.06×10^{-10}
					4	24%	0.01
H							
BT	.70***	.66-.74	0-.92	445,347.62***	1	.07%	0.03×10^{-2}
					2	41%	0.17
					3	0%	0.04×10^{-8}
					4	59%	0.24
T	.68***	.65-.72	.17-.92	499,974.84***	1	.09%	0.01×10^{-3}
					2	41%	0.07×10^{-1}
					3	0%	0.03×10^{-9}
					4	59%	0.01
Raw	.65***	.62-.69	.31-.99	542,959.26***	1	.05%	0.02×10^{-3}
					2	44%	0.01
					3	0%	0.06×10^{-9}
					4	55%	0.02
FD							
BT	.87***	.85-.89	.44-.97	421,266.75***	1	.05%	0.03×10^{-2}
					2	46%	0.26
					3	0	0.04×10^{-5}
					4	54%	0.30
T	.86***	.84-.88	.57-.98	470,003.88***	1	.06%	0
					2	43%	$.06 \times 10^{-1}$
					3	0%	0.02×10^{-9}
					4	57%	0.08×10^{-2}

Raw	.86***	.81-.86	.63-1	551,244.43***	1	.03%	0.03×10^{-4}
					2	40%	0.04×10^{-1}
					3	0%	0.01×10^{-9}
					4	60%	0.06×10^{-1}

*N*ote. Reliability estimates, confidence intervals, and prediction intervals have been back-transformed to the original reliability index scale. Heterogeneity estimates (τ^2) remain in the transformed (or raw) scale. There were 330 entries at level 1/2, 68 at level 3, and 97 at level 4. CI = confidence interval; *ECVs* = Explained Common Variance-specific; *FD* = Factor Determinacy; *H* = Construct Reliability; I^2 = percentage of between-study heterogeneity; ω_s = Coefficient Omega-specific; ω_H = Coefficient Omega Hierarchical-specific; ρ = pooled reliability estimate; PI = prediction interval; *Q* = Cochran's *Q* statistic; τ^2 = between-study variance component.

***$p < .001$

**$p < .01$

*$p < .05$

Table A1.3

Spearman's Rho Correlation Coefficients Among p Factor or (/) specific factor reliability indices

	ECV/ECVs	ω_{HS}/ω_{Hs}	*H*
ω_H/ω_{HS}	0.77/.46		
H	0.43/.60	0.83/.72	
FD	0.45/.39	0.77/.52	0.86/.83

Note. All *r* coefficients were significant at $p < .001$. *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; ω_H = Coefficient Omega Hierarchical.

Table A1.4

Spearman's Rho correlation coefficients among categorical and continuous method variables

Method Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Self v. other	–													
2. Questionnaire v. interview	-0.12	–												
3. Disorder v. symptom level	0.00	0.54	–											
4. Europe, Aus, & SA v. North America	-0.23	-0.15	-0.20	–										
5. PUC	-0.26	-0.16	-0.25	0.09	–									
6. Publication date	0.00	0.14	0.28	-0.22	-0.23	–								
7. Community v. population	-0.08	0.19	0.12	0.15	0.11	0.13	–							
8. Community v. clinical	-0.12	-0.04	-0.01	0.20	0.14	-0.03	-0.22	–						
9. Multiple v. single measures	-0.27	-0.07	0.14	-0.01	-0.06	0.09	-0.06	-0.20	–					
10. Ordinal v. binary indicator	0.20	-0.41	-0.18	0.09	-0.07	-0.11	-0.16	-0.17	0.16	–				
11. WLSMV v. MLR/Bayes	0.13	0.09	0.39	0.05	-0.10	-0.18	-0.06	-0.11	0.19	0.17	–			
12. *N*	-0.18	0.13	-0.10	0.00	0.10	-0.06	0.46	-0.27	-0.04	-0.48	-0.20	–		
13. Age	-0.69	0.20	0.13	0.28	0.32	0.02	0.23	0.18	0.06	-0.31	-0.07	0.24	–	
14. Number of Items	-0.11	-0.31	-0.85	0.19	0.34	-0.34	-0.06	0.03	-0.26	0.00	-0.47	0.18	0.04	–

Note. Aus = Australasia; MLR = Robust Maximum Likelihood estimator; *N* = sample size; PUC = Percentage of Uncontaminated Correlations; SA = South Africa; WLSMV = Weighted Least Squares Means and Variances estimator.

Table A1.5

Multilevel meta-regression coefficients for the method variables predicting p factor reliability indices

Predictor	*β (Ustd)*	*β (Trans)*	*SE*	*z*	*p*	95% CI
ECV						
Intercept	**0.56**	**0.81**	**0.12**	**6.69**	**<.001**	**0.57, 1.05**
PUC (centered)	**0.22**	**0.70**	**0.31**	**2.24**	**0.025**	**0.09, 1.31**
Publication date (log)	-0.02	-0.04	0.06	-0.61	0.542	-0.16, 0.08
Respondent (self *v.* other)	-0.05	-0.10	0.08	-1.23	0.220	-0.27, 0.06
Questionnaire (*v.* interview)	**0.07**	**0.17**	**0.08**	**2.05**	**0.041**	**0.01, 0.34**
Sample (*v.* community)						
Clinical	-0.03	-0.06	0.09	-0.64	0.523	-0.25, 0.13
Population	**0.10**	**0.26**	**0.11**	**2.38**	**0.017**	**0.05, 0.47**
Multiple measures (*v.* single)	-0.06	-0.12	0.09	-1.34	0.181	-0.29, 0.05
Ordinal indicators (*v.* binary)	**0.08**	**0.19**	**0.09**	**2.11**	**0.035**	**0.01, 0.37**
Europe, Aus, & SA (*v.* North America)	-0.02	-0.04	0.09	-0.43	0.667	-0.21, 0.13
MLR/Bayes (vs. WLSMV)	-0.05	-0.11	0.08	-1.36	0.173	-0.27, 0.05
ω_H						
Intercept	**0.70**	**1.19**	**0.18**	**6.61**	**<.001**	**0.84, 1.55**
PUC (centered)	**0.27**	**2.35**	**0.47**	**5.03**	**<.001**	**1.43, 3.26**
Publication date (log)	0.02	0.06	0.09	0.61	0.539	-0.13, 0.24
Respondent (self *v.* other)	0.01	0.04	0.13	0.33	0.741	-0.2, 0.29
Questionnaire (*v.* interview)	**0.09**	**0.34**	**0.13**	**2.70**	**0.007**	**0.09, 0.59**
Sample (*v.* community)						
Clinical	0.02	0.05	0.14	0.36	0.767	-0.23, 0.33
Population	0.06	0.22	0.16	1.37	0.169	-0.09, 0.54
Multiple measures (*v.* single)	**-0.10**	**-0.27**	**0.13**	**-2.09**	**0.036**	**-0.53, -0.02**

Ordinal indicators (*v.* binary)	**0.08**	**0.29**	**0.14**	**2.14**	**0.033**	**0.02, 0.57**
Europe, Aus, & SA (*v.* North America)	0.02	0.06	0.13	0.47	0.639	-0.19, 0.32
MLR/Bayes (vs. WLSMV)	**-0.14**	**-0.38**	**0.12**	**-3.16**	**0.002**	**-0.61, -0.14**
H						
Intercept	**0.88**	**2.14**	**0.30**	**7.26**	**<.001**	**1.56, 2.72**
PUC (centered)	**0.10**	**2.05**	**0.74**	**2.79**	**0.005**	**0.61, 3.50**
Publication date (log)	0.02	0.21	0.15	1.37	0.170	-0.09, 0.5
Respondent (self *v.* other)	0.02	0.21	0.21	1.01	0.310	-0.19, 0.61
Questionnaire (*v.* interview)	0.03	0.28	0.20	1.41	0.158	-0.11, 0.67
Sample (*v.* community)						
Clinical	0.03	0.27	0.23	1.19	0.234	-0.17, 0.72
Population	0.01	0.12	0.26	0.46	0.644	-0.39, 0.63
Multiple measures (*v.* single)	**-0.06**	**-0.42**	**0.21**	**-1.98**	**0.048**	**-0.84, 0**
Ordinal indicators (*v.* binary)	**0.04**	**0.43**	**0.21**	**2.06**	**0.040**	**0.02, 0.83**
Europe, Aus, & SA (*v.* North America)	0.01	0.13	0.21	0.63	0.532	-0.29, 0.55
MLR/Bayes (vs. WLSMV)	**-0.07**	**-0.48**	**0.20**	**-2.44**	**0.015**	**-0.86, -0.09**
FD						
Intercept	**0.92**	**2.55**	**0.30**	**8.50**	**<.0001**	**1.97, 3.14**
PUC (centered)	**0.07**	**1.94**	**0.80**	**2.43**	**0.015**	**0.38, 3.51**
Publication date (log)	0.01	0.10	0.16	0.63	0.531	-0.21, 0.41
Respondent (self *v.* other)	0.01	0.17	0.21	0.79	0.431	-0.25, 0.58
Questionnaire (*v.* interview)	0.02	0.25	0.23	1.12	0.265	-0.19, 0.7
Sample (*v.* community)						
Clinical	**0.03**	**0.43**	**0.24**	**1.81**	**0.070**	**-0.04, 0.89**
Population	0.01	0.12	0.28	0.45	0.656	-0.42, 0.66
Multiple measures (*v.* single)	**-0.04**	**-0.46**	**0.22**	**-2.12**	**0.034**	**-0.88, -0.03**
Ordinal indicators (*v.* binary)	0.02	0.34	0.26	1.31	0.190	-0.17, 0.84
Europe, Aus, & SA (*v.* North America)	0.01	0.11	0.21	0.52	0.604	-0.3, 0.52

MLR/Bayes (vs. WLSMV)	-0.03	-0.31	0.20	-1.58	0.114	-0.7, 0.08

Note. There were 97 entries at level 1/2, and 68 at level 3. Aus = Australasia; *β (Ustd)* = unstandardized regression coefficients, estimated from marginal means; *β (Trans)* = unstandardized regressions coefficients based on Bonett transformed data; CI = confidence interval; *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; MLR = Robust Maximum Likelihood estimator; ω_H = Coefficient Omega Hierarchical; PUC = Percentage of Uncontaminated Correlations; SA = South Africa; *SE* = Standard Error; WLSMV = Weighted Least Squares Means and Variances estimator; *z* = *z* statistic.

Table A1.6

Multilevel meta-regression coefficients for the method variables predicting specific factor reliability indices

Predictor	*β (Ustd)*	*β (Trans)*	*SE*	*z*	*p*	95% CI
ECVs						
Intercept	**0.18**	**0.20**	**0.02**	**8.63**	**<.001**	**0.15, 0.24**
PUC (centered)	**-0.18**	**-0.34**	**0.06**	**-5.66**	**<.001**	**-0.46, -0.22**
Publication date (log)	0.00	0.00	0.01	0.17	0.868	-0.02, 0.03
Respondent (self *v.* other)	0.00	0.00	0.02	-0.28	0.779	-0.04, 0.03
Questionnaire (*v.* interview)	**-0.07**	**-0.08**	**0.02**	**-4.52**	**<.001**	**-0.11, -0.04**
Sample (*v.* community)						
Clinical	-0.02	-0.02	0.02	-1.19	0.232	-0.06, 0.01
Population	**-0.04**	**-0.05**	**0.02**	**-2.51**	**0.012**	**-0.09, -0.01**
Multiple measures (*v.* single)	**0.04**	**0.04**	**0.02**	**2.59**	**0.010**	**0.01, 0.08**
Ordinal indicators (*v.* binary)	0.00	0.00	0.02	-0.20	0.843	-0.04, 0.03
Europe, Aus, & SA (*v.* North America)	0.01	0.01	0.02	0.39	0.694	-0.02, 0.04
MLR/Bayes (vs. WLSMV)	0.02	0.03	0.02	1.64	0.102	-0.01, 0.06
ω_H						
Intercept	**0.48**	**0.65**	**0.11**	**5.86**	**<.001**	**0.43, 0.86**
PUC (centered)	0.01	0.01	0.28	0.05	0.960	-0.54, 0.57
Publication date (log)	0.03	0.07	0.06	1.22	0.222	-0.04, 0.18
Respondent (self *v.* other)	0.01	0.02	0.08	0.23	0.818	-0.13, 0.16
Questionnaire (*v.* interview)	-0.04	-0.07	0.08	-0.91	0.363	-0.23, 0.08
Sample (*v.* community)						
Clinical	-0.01	-0.01	0.08	-0.17	0.864	-0.17, 0.15
Population	**-0.13**	**-0.22**	**0.10**	**-2.27**	**0.023**	**-0.41, -0.03**
Multiple measures (*v.* single)	0.02	0.03	0.08	0.39	0.694	-0.12, 0.18
Ordinal indicators (*v.* binary)	**-0.12**	**-0.20**	**0.09**	**-2.25**	**0.024**	**-0.37, -0.03**

Europe, Aus, & SA (*v.* North America)	0.04	0.07	0.07	1.01	0.311	-0.07, 0.22
MLR/Bayes (vs. WLSMV)	0.05	0.10	0.07	1.50	0.134	-0.03, 0.24
H						
Intercept	**0.74**	**1.34**	**0.23**	**5.86**	**<.001**	**0.89, 1.79**
PUC (centered)	-0.15	-0.46	0.57	-0.81	0.420	-1.57, 0.66
Publication date (log)	0.04	0.17	0.11	1.52	0.128	-0.05, 0.39
Respondent (self *v.* other)	0.04	0.18	0.15	1.16	0.247	-0.12, 0.48
Questionnaire (*v.* interview)	-0.03	-0.10	0.16	-0.67	0.502	-0.41, 0.2
Sample (*v.* community)						
Clinical	-0.01	-0.02	0.17	-0.13	0.899	-0.36, 0.32
Population	-0.11	-0.35	0.20	-1.74	0.081	-0.74, 0.04
Multiple measures (*v.* single)	0.03	0.12	0.16	0.72	0.474	-0.2, 0.43
Ordinal indicators (*v.* binary)	-0.04	-0.15	0.17	-0.85	0.393	-0.48, 0.19
Europe, Aus, & SA (*v.* North America)	0.02	0.09	0.16	0.56	0.578	-0.22, 0.4
MLR/Bayes (vs. WLSMV)	-0.03	-0.11	0.15	-0.79	0.432	-0.4, 0.17
FD						
Intercept	**0.87**	**2.04**	**0.26**	**7.76**	**<.001**	**1.53, 2.56**
PUC (centered)	0.00	-0.03	0.66	-0.04	0.967	-1.32, 1.26
Publication date (log)	0.02	0.17	0.13	1.28	0.201	-0.09, 0.42
Respondent (self *v.* other)	0.02	0.18	0.18	0.99	0.322	-0.17, 0.52
Questionnaire (*v.* interview)	-0.01	-0.10	0.18	-0.55	0.585	-0.45, 0.25
Sample (*v.* community)						
Clinical	0.01	0.09	0.20	0.45	0.652	-0.3, 0.48
Population	-0.06	-0.39	0.23	-1.72	0.086	-0.84, 0.06
Multiple measures (*v.* single)	0.01	0.08	0.18	0.41	0.682	-0.29, 0.44
Ordinal indicators (*v.* binary)	0.00	-0.01	0.20	-0.03	0.975	-0.4, 0.38
Europe, Aus, & SA (*v.* North America)	0.02	0.13	0.18	0.73	0.466	-0.22, 0.49
MLR/Bayes (vs. WLSMV)	-0.02	-0.11	0.17	-0.69	0.493	-0.44, 0.21

Note. There were 330 entries at level 1/2, 97 at level 3, and 68 at level 4. Aus = Australasia; *β (Ustd)* = unstandardized regression coefficients, estimated from marginal means; *β (Trans)* = unstandardized regressions coefficients based on Bonett transformed data; CI = confidence interval; *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; MLR = Robust Maximum Likelihood estimator; ω_H = Coefficient Omega Hierarchical; PUC = Percentage of Uncontaminated Correlations; SA = South Africa; *SE* = Standard Error; WLSMV = Weighted Least Squares Means and Variances estimator; *z* = *z* statistic.

1.3 Meta-regression results after removing outlying and influential cases.

1.3.1 *p* Factor Meta-Regressions

Table A1.7 shows the *p* factor regression coefficients after removing outlying and influential cases.

Explained Common Variance. I removed the six entries for Murray et al. (2016) and MIDAS entry for Levin-Aspenson et al. (2021), as they showed particularly extreme but moderately weighted values in the predicted plots and large studentized residuals (e.g., Murray et al. = -2.4–2.6; Levin-Aspenson et al. = 2.6). The model was significant (QM[10] = 32.44, $p < .001$), explaining 26% of the variance (pseudo-R^2 level 2 = 32%, level 3 = 6%).

There were no major changes in regression coefficients after their removal, but the predicted increase in ECV estimates for population samples compared to community samples was now trending to significance ($\beta = .07$, $z = 1.73$, $p = .083$). Furthermore, the predicted decrease in ECV when combining multiple measures compared to using a single measure was now marginally significant ($\beta = -.07$, $z = 1.92$, $p = .055$). No studies showed concerning DFFITS values.

Omega Hierarchical (ω_H). I removed Miller et al. (2021) as it showed particularly high levels of influence across five predictors (DFFITS = .39-.67) and a large studentized residual ($t = 3.3$). The model was significant ($QM(10) = 118.92$, $p < .001$), explaining 60% of the variance (pseudo-R^2 level 2 = 65%, level 3 = 33%). There were no major differences in regression coefficients or their significance. Furthermore, no studies showed outlying values in the predicted plots.

Construct Reliability (H). I removed both child and mother entries for Martel et al. (2016; DFFITS = .37-1.30) and Vosberg et al. (2021; DFFITS= .43-.57) due to their concerning levels of influence. The model was significant ($QM(10) = 57.11$, $p < .001$), explaining 60% of the variance (pseudo-R^2 level 2 = 65%, level 3 = 33%).

There were no major changes in the 'back-transformed' (i.e. marginal) regression coefficients, but some Bonett-transformed coefficients changed in significance. For example, the marginal increase in *H* for questionnaires compared to interviews was now significant ($\beta = .05$, $z = 2.20$, $p = .028$). Furthermore, the marginal increase in *H* for ordinal compared to binary indicators was no longer significant ($\beta = .03$, $z = .91$, $p = .360$), while the marginal decrease in *H* when combining multiple measures compared to a single measure was now trending ($\beta = -.06$, $z = 1.78$, $p = .075$). There were no studies that showed extreme values in the predicted plots or studentized residuals.

Factor Determinacy (FD). I also removed both child and mother entries for Martel et al. (2016; DFFITS = .37-1.26) and Vosberg et al. (2021; DFFITS = .43-.55). The model was significant ($QM(10) = 41.83$, $p < .001$), explaining 60% of the variance (pseudo-R^2 level 2 = 65%, level 3 = 33%).

There were no major changes in regression coefficients, except the increase in *FD* predicted by an increase in PUC was now trending ($\beta = .07$, $z = 1.81$, $p = .07$), and increase in *FD* for clinical compared to community samples was now marginal ($\beta = .03$, $z = 1.95$, $p = .051$). There were no studies that showed extreme values in the predicted plots or studentized residuals.

1.3.2 Specific Factor Meta-Regressions

Table A1.8 shows the specific factor regression coefficients after removing outlying and influential cases.

Explained Common Variance-Subscale. I removed King et al. (2020) as it showed a high level of influence for nine predictors (DFFITS = |.17|-|.53|) and a large studentized residual (t = 3.9). The model was significant ($QM(10)$ = 125.55, p < .001), explaining 56% of the variance (pseudo-R^2 level 2 = 4%, level 3 = 41%, level 4 = 77%). There were no major changes in regression coefficients.

Omega Hierarchical-Subscale (ω_{HS}). I removed Martel et al. (2017; mothers only) as it showed a high level of influence across five predictors (DFFITS = .|22|-|.43|). The model was marginally significant ($QM(10)$ = 24.96, p = .054), explaining 9% of the variance (pseudo-R^2 level 2 = 0%, level 3 = 24%, level 4 = 35%). The predicted decrease in ω_{HS} associated with ordinal compared to binary indicators strengthened (β = -.18, z = 2.77, p = .006), whilst the predicted increase in ω_{HS} associated with MLR/Bayes estimators compared to WLSMV became significant (β = .07, z = 2.17, p = .03). There were no studies that showed extreme values in the predicted plots. Studentized residuals could not be computed for all entries.

Construct Reliability (H). No studies showed influential values or studentized residuals that were of concern. The model was not significant ($QM(10)$ = 12.87, p = .231) and explained 2% of the variance (pseudo-R^2 level 2 = 0.4%, level 3 = 4%, level 4 = 4%).

Factor Determinacy (FD). I removed Brandes et al. (2019) for its high levels of influence across seven predictors (DFFITS = .|17|-|.46|) and Niarchou et al.

(2017) for its large studentized residuals (ts = 2.4-3), but there were no major changes in regression coefficients. The model was not significant ($QM(10) = 12.13$, $p = .276$), explaining 2% of the variance (pseudo-R^2 level 2 = 0%, level 3 = 4%, level 4 = 3%).

Table A1.7

Multilevel meta-regression coefficients for the method variables predicting p factor reliability indices after removing outlying and/or influential cases

Predictor	*β (Ustd)*	*β (Trans)*	*SE*	*z*	*p*	95% CI
ECV						
Intercept	**0.57**	**0.84**	**0.11**	**7.32**	**<.001**	**0.61, 1.06**
PUC (centered)	**0.20**	**0.63**	**0.30**	**2.08**	**0.037**	**0.04, 1.23**
Publication date (log)	-0.01	-0.03	0.06	-0.57	0.571	-0.15, 0.08
Respondent (self *v.* other)	-0.06	-0.13	0.08	-1.55	0.120	-0.28, 0.03
Questionnaire (*v.* interview)	**0.07**	**0.18**	**0.08**	**2.20**	**0.028**	**0.02, 0.34**
Sample (*v.* community)						
Clinical	-0.04	-0.08	0.10	-0.87	0.387	-0.27, 0.11
Population	0.07	0.18	0.11	1.73	0.084	-0.02, 0.39
Multiple measures (*v.* single)	-0.08	-0.16	0.08	-1.92	0.055	-0.33, 0
Ordinal indicators (*v.* binary)	**0.07**	**0.18**	**0.09**	**1.97**	**0.049**	**0, 0.36**
Europe, Aus, & SA (*v.* North America)	0.01	0.02	0.08	0.21	0.834	-0.15, 0.18
MLR/Bayes (vs. WLSMV)	-0.03	-0.07	0.08	-0.95	0.341	-0.23, 0.08
ω_H						
Intercept	**0.68**	**1.14**	**0.17**	**6.80**	**<.001**	**0.81, 1.46**
PUC (centered)	**0.29**	**2.38**	**0.43**	**5.51**	**<.001**	**1.53, 3.23**
Publication date (log)	0.01	0.05	0.09	0.55	0.582	-0.12, 0.22
Respondent (self *v.* other)	0.03	0.10	0.12	0.84	0.399	-0.13, 0.33
Questionnaire (*v.* interview)	**0.09**	**0.33**	**0.12**	**2.75**	**0.006**	**0.09, 0.56**
Sample (*v.* community)						
Clinical	0.02	0.06	0.13	0.48	0.633	-0.19, 0.32
Population	0.06	0.22	0.15	1.46	0.146	-0.08, 0.51
Multiple measures (*v.* single)	**-0.13**	**-0.35**	**0.12**	**-2.84**	**0.005**	**-0.58, -0.11**

Ordinal indicators (*v.* binary)	**0.08**	**0.28**	**0.13**	**2.17**	**0.030**	**0.03, 0.53**
Europe, Aus, & SA (*v.* North America)	0.03	0.11	0.12	0.90	0.368	-0.13, 0.34
MLR/Bayes (vs. WLSMV)	**-0.11**	**-0.31**	**0.11**	**-2.71**	**0.007**	**-0.53, -0.08**
H						
Intercept	**0.87**	**2.07**	**0.31**	**6.60**	**<.001**	**1.45, 2.68**
PUC (centered)	**0.11**	**1.97**	**0.79**	**2.50**	**0.012**	**0.43, 3.51**
Publication date (log)	0.03	0.26	0.16	1.62	0.105	-0.05, 0.57
Respondent (self *v.* other)	0.02	0.16	0.21	0.80	0.426	-0.24, 0.57
Questionnaire (*v.* interview)	**0.05**	**0.51**	**0.23**	**2.20**	**0.028**	**0.06, 0.96**
Sample (*v.* community)						
Clinical	0.03	0.28	0.24	1.19	0.234	-0.18, 0.74
Population	0.02	0.17	0.27	0.63	0.527	-0.36, 0.7
Multiple measures (*v.* single)	-0.06	-0.39	0.22	-1.78	0.075	-0.82, 0.04
Ordinal indicators (*v.* binary)	0.03	0.25	0.28	0.91	0.360	**-0.29, 0.8**
Europe, Aus, & SA (*v.* North America)	0.02	0.16	0.22	0.72	0.471	-0.27, 0.58
MLR/Bayes (vs. WLSMV)	**-0.06**	**-0.41**	**0.20**	**-2.00**	**0.046**	**-0.8, -0.01**
FD						
Intercept	**0.91**	**2.41**	**0.32**	**7.54**	**<.001**	**1.78, 3.03**
PUC (centered)	**0.07**	**1.54**	**0.85**	**1.81**	**0.070**	**-0.13, 3.2**
Publication date (log)	0.01	0.18	0.17	1.09	0.276	-0.15, 0.51
Respondent (self *v.* other)	0.01	0.17	0.21	0.78	0.434	-0.25, 0.59
Questionnaire (*v.* interview)	**0.02**	**0.29**	**0.24**	**1.22**	**0.222**	**-0.18, 0.76**
Sample (*v.* community)						
Clinical	**0.03**	**0.48**	**0.25**	**1.95**	**0.051**	**0, 0.97**
Population	0.02	0.21	0.29	0.73	0.467	-0.36, 0.78
Multiple measures (*v.* single)	**-0.05**	**-0.45**	**0.22**	**-2.00**	**0.046**	**-0.88, -0.01**
Ordinal indicators (*v.* binary)	**0.03**	**0.43**	**0.29**	**1.47**	**0.141**	**-0.14, 1**
Europe, Aus, & SA (*v.* North America)	0.01	0.11	0.21	0.51	0.608	-0.31, 0.53

MLR/Bayes (vs. WLSMV)	**-0.03**	**-0.30**	**0.21**	**-1.48**	**0.139**	**-0.71, 0.10**

Note. The number of entries varied depending on the number of studies removed. I removed seven entries for *ECV* (Murray et al., 2016 and Levin-Aspenson et al., 2021), one entry for ω_H (Miller et al., 2021), and three entries for both *H* and *FD* (Martel et al., 2017 and Vosberg et al., 2021). Aus = Australasia; *β (Ustd)* = unstandardized regression coefficients, estimated from marginal means; *β (Trans)* = unstandardized regressions coefficients based on Bonett transformed data; CI = confidence interval; *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; MLR = Robust Maximum Likelihood estimator; ω_H = Coefficient Omega Hierarchical; PUC = Percentage of Uncontaminated Correlations; SA = South Africa; *SE* = Standard Error; WLSMV = Weighted Least Squares Means and Variances estimator; *z* = *z* statistic.

Table A1.8

Multilevel meta-regression coefficients for the method variables predicting specific factor reliability indices after removing outlying and/or influential cases

Predictor	*β (Ustd)*	*β (Trans)*	*SE*	*z*	*p*	95% CI
ECVs						
Intercept	**0.18**	**0.20**	**0.02**	**8.54**	**<.001**	**0.15, 0.25**
PUC (centered)	**-0.18**	**-0.34**	**0.06**	**-5.59**	**<.001**	**-0.46, -0.22**
Publication date (log)	0.00	0.00	0.01	0.26	0.797	-0.02, 0.03
Respondent (self *v.* other)	0.00	-0.01	0.02	-0.36	0.718	-0.04, 0.03
Questionnaire (*v.* interview)	**-0.06**	**-0.07**	**0.02**	**-4.33**	**<.001**	**-0.11, -0.04**
Sample (*v.* community)						
Clinical	-0.02	-0.02	0.02	-1.21	0.225	-0.06, 0.01
Population	**-0.04**	**-0.05**	**0.02**	**-2.50**	**0.012**	**-0.09, -0.01**
Multiple measures (*v.* single)	**0.03**	**0.04**	**0.02**	**2.50**	**0.012**	**0.01, 0.08**
Ordinal indicators (*v.* binary)	0.00	-0.01	0.02	-0.28	0.777	-0.04, 0.03
Europe, Aus, & SA (*v.* North America)	0.01	0.01	0.02	0.50	0.616	-0.02, 0.04
MLR/Bayes (vs. WLSMV)	0.02	0.02	0.02	1.51	0.130	-0.01, 0.05
ω_H						
Intercept	**0.49**	**0.67**	**0.11**	**6.32**	**<.001**	**0.46, 0.88**
PUC (centered)	0.02	0.04	0.27	0.13	0.897	-0.5, 0.57
Publication date (log)	0.03	0.06	0.05	1.16	0.246	-0.04, 0.17
Respondent (self *v.* other)	0.01	0.02	0.07	0.26	0.799	-0.12, 0.16
Questionnaire (*v.* interview)	-0.01	-0.01	0.08	-0.14	0.888	-0.17, 0.15
Sample (*v.* community)						
Clinical	-0.03	-0.06	0.08	-0.74	0.457	-0.22, 0.1
Population	**-0.15**	**-0.26**	**0.09**	**-2.77**	**0.006**	**-0.44, -0.08**
Multiple measures (*v.* single)	0.00	0.01	0.07	0.07	0.942	-0.14, 0.15

Ordinal indicators (*v.* binary)	-0.18	-0.30	0.09	-3.16	0.002	-0.48, -0.11
Europe, Aus, & SA (*v.* North America)	0.06	0.12	0.07	1.60	0.110	-0.03, 0.26
MLR/Bayes (vs. WLSMV)	**0.07**	**0.15**	**0.07**	**2.17**	**0.030**	**0.01, 0.28**
FD						
Intercept	**0.87**	**2.01**	**0.27**	**7.47**	**<.001**	**1.48, 2.53**
PUC (centered)	-0.02	-0.13	0.68	-0.20	0.842	-1.46, 1.19
Publication date (log)	0.02	0.17	0.13	1.32	0.186	-0.08, 0.43
Respondent (self *v.* other)	0.02	0.20	0.18	1.11	0.266	-0.15, 0.56
Questionnaire (*v.* interview)	-0.01	-0.06	0.19	-0.34	0.737	-0.43, 0.3
Sample (*v.* community)						
Clinical	0.01	0.09	0.20	0.45	0.654	-0.3, 0.48
Population	-0.06	-0.38	0.23	-1.65	0.099	-0.83, 0.07
Multiple measures (*v.* single)	0.01	0.05	0.19	0.25	0.801	-0.32, 0.42
Ordinal indicators (*v.* binary)	0.00	-0.01	0.20	-0.07	0.941	-0.41, 0.38
Europe, Aus, & SA (*v.* North America)	0.02	0.14	0.18	0.77	0.441	-0.22, 0.5
MLR/Bayes (vs. WLSMV)	-0.02	-0.12	0.17	-0.73	0.468	-0.45, 0.21

Note. The number of entries varied depending on the number of studies removed. I removed one entry for *ECV* (King et al., 2020), one entry for ω_H (Martel et al., 2017), no entries for *H* (not shown), and two entries for *FD* (Brandes et al., 2019 and Niarchou et al., 2017). Aus = Australasia; *β (Ustd)* = unstandardized regression coefficients, estimated from marginal means; *β (Trans)* = unstandardized regressions coefficients based on Bonett transformed data; CI = confidence interval; *ECV* = Explained Common Variance; *FD* = Factor Determinacy; *H* = Construct Reliability; MLR = Robust Maximum Likelihood estimator; ω_H = Coefficient Omega Hierarchical; PUC = Percentage of Uncontaminated Correlations; SA = South Africa; *SE* = Standard Error; WLSMV = Weighted Least Squares Means and Variances estimator; *z* = *z* statistic.

Appendix 2. Empirical Paper

2.1 Moderated Mediation Model Controlling for Covariates

I replicated the moderated mediation model in the main analysis but regressed socioeconomic status (SES), stressful life events (SLEs), family obligation, the interaction term between family obligation and SLEs, and general and specific psychopathology factor scores onto child age, gender, and racial background.

As seen in Table A1, children from Black/African American, Native American, and Hispanic families showed lower SES scores than children from white families, whereas children from Asian families showed slightly higher SES scores than white children. Older children showed slightly higher SES scores and children from Asian families showed slightly lower SLE scores. Children who were older, male (vs. female), and from Black/African American (vs. white) showed higher family obligation scores, as did Hispanic children though the result was marginal. Older children also showed a slightly stronger interaction term between SLEs and family obligation scores. *p* factor scores were slightly higher in males compared to females. Furthermore, specific internalizing factor scores were lower in males compared to females and Black/African American children compared to white children, but were marginally higher in Native American children. Specific externalizing factor scores were slightly higher in males and transgender children (marginal) than females, as well as Black African/American children compared to white children. Lastly, specific attention factor scores were slightly lower in Asian children and marginally lower in transgender children.

Table A2.1

Standardized regression coefficients for the child covariates predicting all other variables in the in the moderated mediation model

Factor/Covariate	*B*	BS 95% CI	*z*	*p*
p factor				
Male (vs. female)	**0.02**	**0.01, 0.04**	**2.28**	**.023**
Trans (vs. female)	0.00	-0.02, 0.02	0.02	.987
Child Age	0.00	-0.02, 0.02	0.24	.811
Black (vs. white)	-0.01	-0.03, 0.01	-1.00	.316
Native American (vs. white)	0.00	-0.01, 0.01	-0.31	.757
Asian (vs. white)	0.00	-0.02, 0.02	0.06	.949
Hispanic (vs. white)	0.01	-0.01, 0.02	1.10	.273
Specific Internalizing				
Male (vs. female)	**-0.19**	**-0.2, -0.17**	**-20.05**	**< .001**
Trans (vs. female)	-0.01	-0.03, 0.01	-1.21	.226
Child Age	0.01	-0.01, 0.03	1.01	.315
Black (vs. white)	**-0.05**	**-0.08, -0.03**	**-5.14**	**< .001**
Native American (vs. white)	0.02	0, 0.03	1.93	.054
Asian (vs. white)	0.01	-0.01, 0.03	0.79	.431
Hispanic (vs. white)	0.00	-0.02, 0.01	-0.25	.807
Specific Externalizing				
Male (vs. female)	**0.02**	**0.01, 0.04**	**2.28**	**.023**
Trans (vs. female)	0.02	0, 0.05	1.65	.099
Child Age	0.00	-0.02, 0.02	0.44	.659
Black (vs. white)	**0.03**	**0.01, 0.05**	**2.34**	**.019**
Native American (vs. white)	0.01	-0.01, 0.03	1.42	.155
Asian (vs. white)	0.01	-0.01, 0.03	0.83	.408
Hispanic (vs. white)	0.01	-0.01, 0.03	0.70	.484
Specific Attention				
Male (vs. female)	0.01	-0.01, 0.03	0.97	.331
Trans (vs. female)	-0.01	-0.03, 0	-1.84	.065
Child Age	0.01	-0.01, 0.03	1.02	.309
Black (vs. white)	0.00	-0.02, 0.02	0.18	.861
Native American (vs. white)	-0.01	-0.02, 0.02	-0.66	.511
Asian (vs. white)	**-0.04**	**-0.06, -0.02**	**-4.06**	**< .001**
Hispanic (vs. white)	-0.01	-0.03, 0.01	-0.78	.438
SES				
Male (vs. female)	0.00	-0.01, 0.02	0.50	.617
Trans (vs. female)	0.00	-0.02, 0.02	0.17	.865
Child Age	**0.05**	**0.03, 0.07**	**5.19**	**< .001**
Black (vs. white)	**-0.46**	**-0.48, -0.44**	**-48.19**	**< .001**
Native American (vs. white)	**-0.01**	**-0.03, 0**	**-2.32**	**.02**
Asian (vs. white)	**0.05**	**0.03, 0.06**	**5.73**	**< .001**
Hispanic (vs. white)	**-0.03**	**-0.06, -0.01**	**-2.98**	**.003**
SLEs				

Male (vs. female)	0.00	-0.02, 0.02	-0.11	0.91
Trans (vs. female)	0.01	-0.01, 0.03	0.57	0.567
Child Age	0.00	-0.02, 0.02	0.27	0.785
Black (vs. white)	-0.01	-0.03, 0.02	-0.59	0.555
Native American (vs. white)	-0.01	-0.01, 0.01	-1.07	0.284
Asian (vs. white)	**-0.04**	**-0.06, -0.03**	**-5.17**	**< .001**
Hispanic (vs. white)	0.02	-0.01, 0.05	1.10	0.27
Family Obligation				
Male (vs. female)	**0.09**	**0.08, 0.11**	**9.66**	**< .001**
Trans (vs. female)	0.00	-0.02, 0.03	0.26	.792
Child Age	**-0.05**	**-0.07, -0.03**	**-5.01**	**< .001**
Black (vs. white)	**0.14**	**0.12, 0.16**	**13.55**	**< .001**
Native American (vs. white)	0.01	-0.01, 0.04	1.07	.285
Asian (vs. white)	-0.01	-0.03, 0.01	-1.07	.285
Hispanic (vs. white)	0.02	0, 0.04	1.77	.077
Family Obligation*SLEs				
Male (vs. female)	0.01	-0.01, 0.03	1.38	.166
Trans (vs. female)	-0.02	-0.06, 0.01	-0.83	.407
Child Age	**-0.03**	**-0.05, -0.003**	**-2.32**	**.02**
Black (vs. white)	0.01	-0.01, 0.03	0.82	.415
Native American (vs. white)	0.00	0, 0.01	1.03	.305
Asian (vs. white)	0.01	-0.01, 0.02	1.09	.277
Hispanic (vs. white)	0.01	-0.01, 0.05	0.92	.356

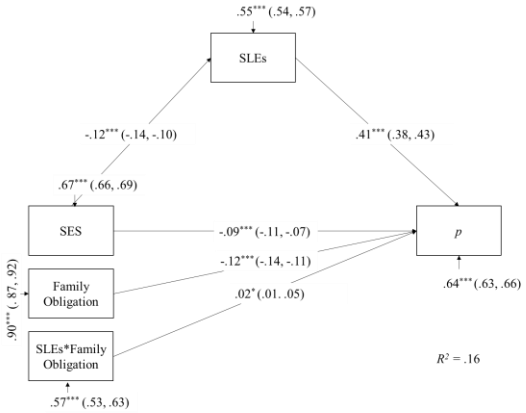
Note. BS = Bootstrapped. SES = Socioeconomic Status; SLEs = Stressful Life Events. Significant results in bold.

Figure A2.1 shows the moderated mediation path coefficients corrected for age, gender and ethnicity. The direct and indirect path coefficients were similar in magnitude and significance to the main analysis, except for specific internalizing scores, where the strength of family obligation and SES's predictions halved but remained significant and not significant, respectively. The indirect effects of SES on the general and specific psychopathology factors via SLEs and the indices of moderated mediation were similar between the models with and without covariates. Total effects were similar between the models, except for the total effect of SES and SLEs on specific internalizing scores, which was slightly larger in the model with covariates (see Table A2).

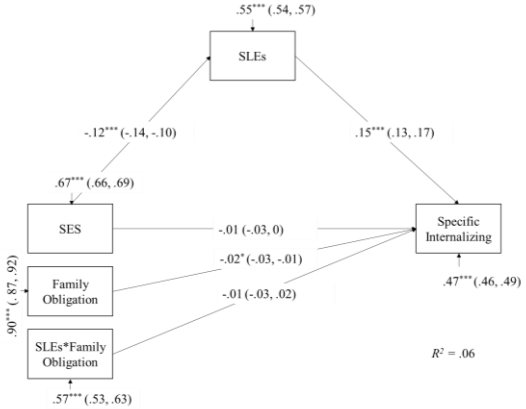
Figure A2.1

Conditional moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the p factor (A), specific internalizing factor (B), specific externalizing factor (C), and specific attention factor (D), after controlling each variable for child age, gender, and racial background

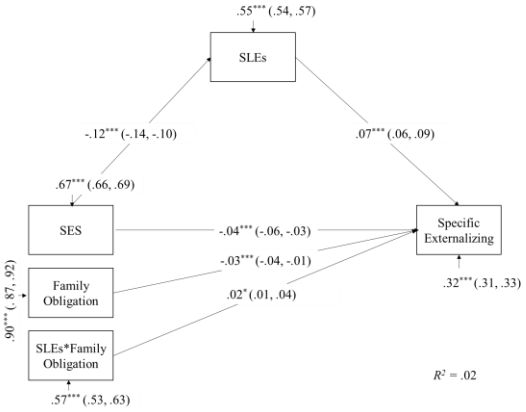
A)



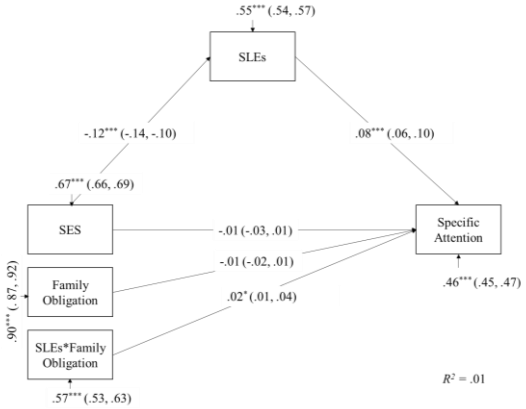
B)



C)



D)



$^{***}p \leq .001$; $^{**}p \leq .01$; $^{*}p \leq .05$; $^{\dagger}p \leq .10$

Table A2.2

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs) after controlling for child age, gender, and racial background

	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
p	-.14***	-.16, -.12	-.05***	-.06, -.04	35%	-.003*	-.005, -.001
Internalizing	-.03*	-.05, -.01	-.02***	-.02, -.01	87%	.001	-.002, .003
Externalizing	-.06***	-.07, -.04	-.01***	-.01, -.01	14%	-.002*	-.005, -.001
Attention	-.02*	-.04, -.01	-.01***	-.01, -.01	41%	-.003*	-.005, -.001

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.

2.2 Controlling for Pre-Existing Psychopathology

I replicated the moderated mediation model in the main analysis whilst regressing SES, SLEs, family obligation, the interaction term between family obligation and SLEs onto general and specific psychopathology factor scores estimated the year before (i.e. *t*-1 factors from Release 2.0). Factors were also autoregressed onto themselves from the year before (e.g., *p* factor scores estimated at Release 3.0 were regressed on *p* factor scores estimated at Release 2.0).

There was partial support for metric invariance between bifactor models estimated at Release 2.0 and 3.0 (e.g., holding factor loadings constant across time-points was associated with an improvement in ΔBIC = 248.23, ΔBIC*n* = 140.17, ΔCFI = .04, and ΔTLI = .02, but a worsening in ΔAIC = -58.43 and no substantial change in ΔRMSEA = .004). Similarly, there was partial support for scalar invariance (e.g., holding both factor loadings and item response thresholds resulted in an improvement in ΔBIC = 13.58 and no substantial changes in ΔCFI = .006, ΔTLI = -.001, and ΔRMSEA = 0, but not ΔAIC = -329.16 or ΔBIC*n* = -107.19). Factor loadings for the bifactor model estimated a year earlier are shown in Table A3.

Table A2.3

Standardized factor loadings for the bifactor model of the Brief Problem Monitor child-reported items at Release 2.0 (t-1)

Item	*p*	Ext	Int	Att
Acts young	0.57			0.03
Argues	0.61	0.43		
Can't finish things	0.59			0.14
Can't concentrate	0.67			0.53
Can't sit still	0.58			0.39
Destroys things	0.60	0.26		
Disobedient (home)	0.58	0.28		
Disobedient (school)	0.60	0.24		
Worthless	0.63		0.46	
Impulsive	0.71			0.01
Fearful	0.54		0.51	
Guilty	0.57		0.45	
Embarrassed	0.51		0.43	
Distracted	0.70			0.49
Stubborn	0.53	0.22		
Temper	0.63	0.43		
Threatens	0.62	0.37		
Unhappy	0.60		0.40	
Worries	0.53		0.55	
M	0.60	0.32	0.47	0.26
SD	0.06	0.09	0.06	0.23
ECV	0.71	0.08	0.14	0.07
ω	0.94	0.86	0.87	0.86
ω_H	0.84	0.19	0.36	0.12
Rel. ω	0.89	0.22	0.41	0.14
H	0.92	0.47	0.63	0.47
FD	0.93	0.68	0.80	0.75

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; ω = Omega; ω_H = Omega Hierarchical.

Higher *t*-1 *p* factor scores and specific internalizing, externalizing, and attention scores predicted lower SES scores and higher SLE scores (except specific externalizing; see Table A4). Lower *t*-1 *p* factor scores and higher *t*-1 specific internalizing and attention scores predicted higher family obligation scores. All *t*-1 psychopathology factors moderately predicted their respective factor scores.

Table A2.4

*Standardized regression coefficients for t-1 general and specific psychopathology factors predicting SES, SLEs, Family Obligation and Family Obligation*SLEs interaction term in the moderated mediation model*

Factor/Covariate	*B*	BS 95% CI	*z*	*p*
SES				
p (*t*-1)	-0.08	-0.11, -0.06	-6.89	**< .001**
Specific attention (*t*-1)	-0.09	-0.12, -0.06	-5.57	**< .001**
Specific externalizing (*t*-1)	-0.18	-0.22, -0.14	-8.65	**< .001**
Specific internalizing (*t*-1)	-0.07	-0.1, -0.04	-4.83	**< .001**
SLEs				
p (*t*-1)	0.20	0.18, 0.22	20.98	**< .001**
Specific attention (*t*-1)	0.04	0.02, 0.07	3.3	**< .001**
Specific externalizing (*t*-1)	0.03	-0.01, 0.06	1.63	0.104
Specific internalizing (*t*-1)	0.07	0.05, 0.09	5.68	**< .001**
Family Obligation				
p (*t*-1)	-0.13	-0.16, -0.11	-10.17	**< .001**
Specific attention (*t*-1)	0.07	0.04, 0.1	4.13	**< .001**
Specific externalizing (*t*-1)	0.03	-0.01, 0.08	1.56	0.118
Specific internalizing (*t*-1)	**0.03**	**0, 0.06**	**2.08**	**0.037**
Family Obligation*SLEs				
p (*t*-1)	-0.01	-0.03, 0.01	-0.8	0.422
Specific attention (*t*-1)	0.01	-0.01, 0.04	0.9	0.366
Specific externalizing (*t*-1)	0	-0.03, 0.04	0.23	0.82
Specific internalizing (*t*-1)	-0.01	-0.04, 0.02	-0.82	0.415
Autoregressions				
p on *p* *t*-1	0.56	**0.55, 0.58**	68	**< .001**
Int on int. *t*-1	0.47	**0.45, 0.5**	44.99	**< .001**
Ext on ext *t*-1	0.43	**0.41, 0.45**	41.57	**< .001**
Att on att *t*-1	0.43	**0.41, 0.45**	43.64	**< .001**

Note. Att = specific attention factor; BS = Bootstrapped. SES = Socioeconomic Status; ext = specific externalizing factor; int = specific internalizing factor; SLEs = Stressful Life Events. Significant results in bold.

Figure A2.2 shows the moderated mediation path coefficients after controlling for *t*-1 *p* factor and specific factor psychopathology scores. Direct and indirect paths were weaker, particularly those predicting psychopathology factors, which is expected given the moderately strong autoregressive coefficients. Nonetheless, the direction and significance of most path coefficients, indirect effects, and indices of moderated mediation remained similar to the main analysis model (see Table A5). However, SLE scores went from negatively to positively predicting specific internalizing factor scores. Also, higher SES scores now marginally predicted higher specific internalizing factor scores but the prediction was still weak (B = .01, BS 95% CI [-.01, .03], p = .065). Lastly, higher SES scores no longer marginally predicted lower specific attention scores (B = 0, BS 95% CI [-.02, .01], p = .749; see Table A5).

Table A2.5

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs) after controlling for bifactor dimensions estimated a year before (at Release 2.0)

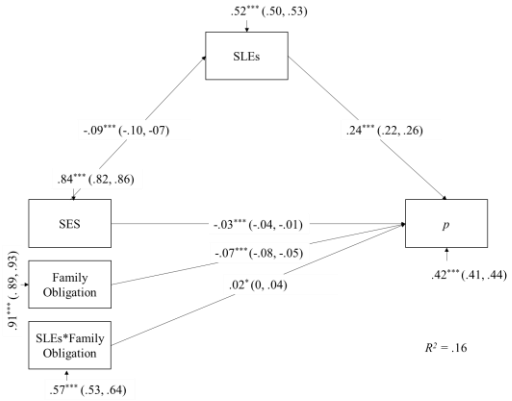
	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
p	-.05***	-.06, -03	-.02***	-.03, -.02	46%	-.002*	-.003, -.001
Internalizing	0	-.01, .02	-.009***	-.01, .01	N/A[a]	0	-.001, .002
Externalizing	-.03***	-.04, -.02	-.005***	-.007, -.004	17%	-.002*	-.003, -.001
Attention	-.006	-.02, .01	-.004***	-.006, -.003	67%	-.002*	-.003, -.001

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.
[a]The total effect was virtually zero so a ratio could not be computed.

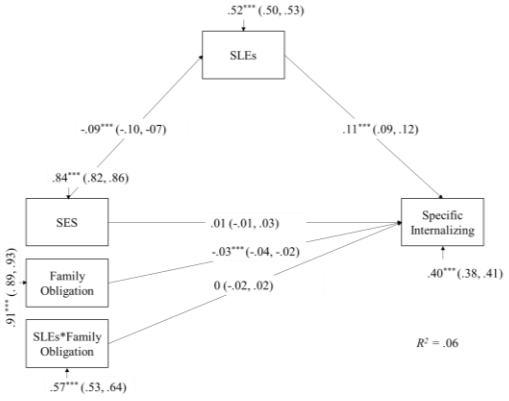
Figure A2.2

Conditional moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the p factor (A), specific internalizing factor (B), specific externalizing factor (C), and specific attention factor (D), after controlling each variable for the p factor, specific internalizing, externalizing, and attention factors estimated at Release 2.0 (t-1)

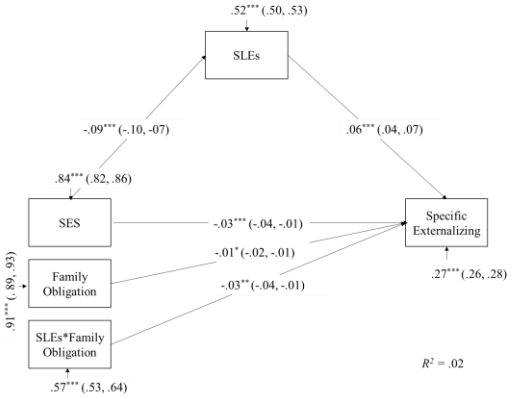
A)



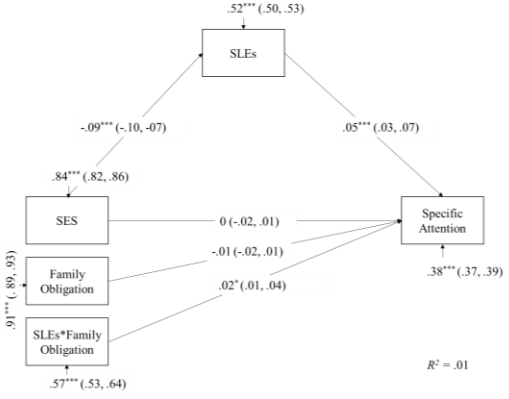
B)



C)



D)



$^{***}p \leq .001$; $^{**}p \leq .01$; $^{*}p \leq .05$; $^{\dagger}p \leq .10$

2.3 Removing SLE Indicators Related to SES

Figure A2.3 presents a moderated mediation model using SLE factor scores estimated after removing SLE indicators that overlapped with SES (e.g., 'Negative change in parent's financial situation', 'Mother/father lost their job', and 'One parent was away from home more often'). Path coefficients, indirect effects, and indices of moderated mediation were similar across models (see Table A6).

Table A2.6

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs) after controlling removing SLE items that overlapped with SES indicators

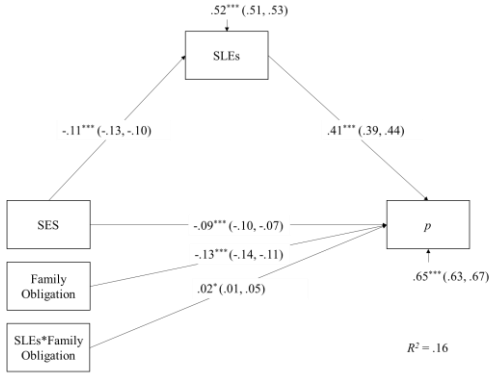
	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
p	-.13***	-.15, -.11	-.05***	-.05, -.04	36%	-.003*	-.005, -.001
Internalizing	-.02*	-.03, -.01	-.02***	-.02, -.01	88%	.001	-.001, .003
Externalizing	-.06***	-.07, -.05	-.01***	-.01, -.01	14%	-.002*	-.004, -.001
Attention	-.02*	-.04, -.01	-.01***	-.01, -.01	45%	-.003*	-.005, -.001

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.

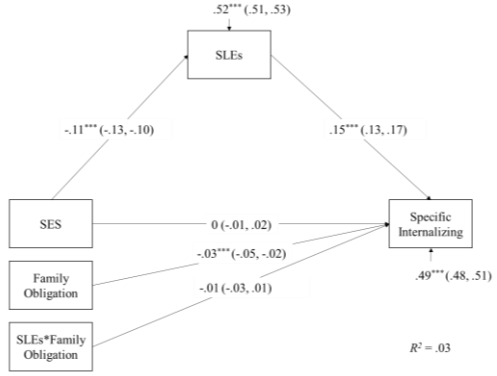
Figure A2.3

Conditional moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the p factor (A), specific internalizing factor (B), specific externalizing factor (C), and specific attention factor (D), after removing SLE items that overlapped with SES indicators

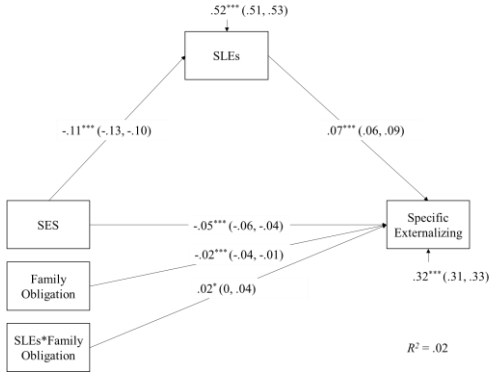
A)



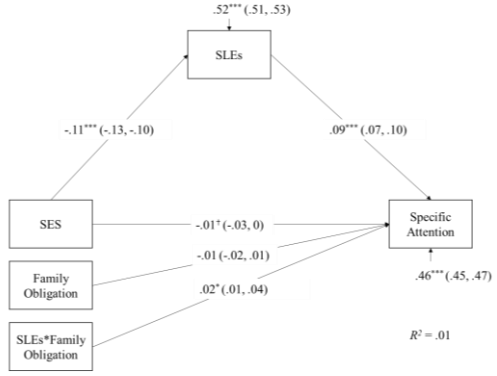
B)



C)



D)



***$p \leq .001$; **$p \leq .01$; *$p \leq .05$; †$p \leq .10$

2.4 SES as Separate but Parallel Predictors

Table A7 shows the direct and indirect paths from the correlated SES indicators to the *p* factor and specific internalizing, externalizing, and attention factors. Regarding the direct paths from each SES indicator to psychopathology factor scores (c'), higher household income weakly predicted lower *p* factor scores and specific attention scores. Furthermore, the presence of a partner in the family household weakly predicted higher *p* factor and specific internalizing factor scores. More social and economic neighbourhood opportunities weakly predicted lower specific externalizing scores and marginally predicted lower *p* factor scores. Higher poverty scores weakly predicted higher *p* factor and specific attention scores. Lastly, higher parental education weakly predicted lower specific attention scores. SES predictors showed small to moderate positive correlations (see Table A8).

Table A2.7

Standardized paths from each SES indicator predicting the p factor and specific internalizing, externalizing, and attention factors directly and indirectly via SLEs

Path	*B*	BS 95% CI	*z*	*p*
SES → *p* (*c′*)				
Family Household Income	**-0.03**	**-0.06, 0**	**-2.05**	**0.04**
Educational Attainment	-0.01	-0.04, 0.01	-1.14	0.255
Single-partner household	**0.03**	**0.01, 0.05**	**2.37**	**0.018**
COI Education	-0.02	-0.05, 0.01	-1.53	0.127
COI Health and Education	-0.02	-0.05, 0.01	-1.25	0.211
COI Social and Economic	-0.03	-0.06, 0.01	-1.69	0.092
Poverty score	**-0.03**	**-0.05, -0.01**	**-2.99**	**0.003**
SES → Specific Internalizing (*c′*)				
Family Household Income	0	-0.03, 0.03	-0.04	0.971
Educational Attainment	0.01	-0.02, 0.03	0.38	0.702
Single-partner household	**0.03**	**0.01, 0.06**	**2.99**	**0.003**
COI Education	0	-0.03, 0.03	0.25	0.803
COI Health and Education	0	-0.03, 0.03	-0.24	0.809
COI Social and Economic	-0.01	-0.05, 0.02	-0.63	0.527
Poverty score	0	-0.02, 0.02	-0.15	0.883
SES → Specific Externalizing (*c′*)				
Family Household Income	-0.02	-0.05, 0.01	-1.24	0.213
Educational Attainment	-0.02	-0.04, 0.01	-1.36	0.174
Single-partner household	-0.01	-0.03, 0.02	-0.39	0.696
COI Education	0	-0.03, 0.03	-0.02	0.983
COI Health and Education	-0.02	-0.05, 0.02	-1.07	0.285
COI Social and Economic	**-0.04**	**-0.08, -0.01**	**-2.38**	**0.017**
Poverty score	-0.01	-0.03, 0.01	-0.74	0.458
SES → Specific Attention (*c′*)				
Family Household Income	**0.03**	**0, 0.07**	**2.14**	**0.032**
Educational Attainment	**-0.03**	**-0.06, -0.01**	**-2.53**	**0.011**
Single-partner household	0	-0.03, 0.02	-0.33	0.742
COI Education	0	-0.03, 0.03	-0.04	0.966
COI Health and Education	-0.02	-0.05, 0.01	-1.15	0.25
COI Social and Economic	0.01	-0.03, 0.04	0.32	0.747
Poverty score	**-0.04**	**-0.06, -0.02**	**-3.37**	**< .001**
SES → SLEs (a_1)				
Family Household Income	**-0.07**	**-0.11, -0.04**	**-4.48**	**< .001**
Educational Attainment	-0.02	-0.04, 0.01	-1.3	0.194
Single-partner household	**-0.14**	**-0.16, -0.12**	**-11.85**	**< .001**
COI Education	**-0.04**	**-0.07, -0.01**	**-2.56**	**0.011**
COI Health and Education	0.01	-0.02, 0.04	0.56	0.576
COI Social and Economic	0.02	-0.01, 0.06	1.2	0.23
Poverty score	**-0.08**	**-0.11, -0.06**	**-7.06**	**< .001**
SLEs → psychopathology (b_1)				

p	**0.4**	**0.38, 0.43**	**36.55**	**< .001**
Specific Internalizing	**0.15**	**0.13, 0.17**	**14.97**	**< .001**
Specific Externalizing	**0.07**	**0.06, 0.09**	**8.78**	**< .001**
Specific Attention	**0.08**	**0.06, 0.1**	**8.36**	**< .001**
FO → psychopathology (b_2)				
p	**-0.13**	**-0.14, -0.11**	**-14.16**	**< .001**
Specific Internalizing	**-0.03**	**-0.05, -0.02**	**-4.36**	**< .001**
Specific Externalizing	**-0.02**	**-0.04, -0.01**	**-4.04**	**< .001**
Specific Attention	-0.01	-0.02, 0.01	-1.15	0.248
FO*SLE → psychopathology (b_3)				
p	**0.02**	**0, 0.05**	**2.18**	**0.029**
Specific Internalizing	-0.01	-0.03, 0.01	-0.63	0.53
Specific Externalizing	**0.02**	**0, 0.04**	**2.4**	**0.016**
Specific Attention	**0.02**	**0.01, 0.04**	**2.47**	**0.014**

Note. BS = Bootstrapped. COI = Child Opportunity Index; FO = Family Obligation; SES = Socioeconomic Status; SLEs = Stressful Life Events. Significant results in bold.

Table A2.8

Correlation coefficients for the SES indicators

SES Indicator	1.	2.	3.	4.	5.	6.	7.
1. Family Household Income	–						
2. Educational Attainment	.58***	–					
3. Single-partner household	.45***	.19***	–				
4. COI Education	.49***	.42***	.19***	–			
5. COI Health and Education	.49***	.37***	.22***	.61***	–		
6. COI Social and Economic	.58***	.46***	.25***	.71***	.73***	–	
7. Poverty score	.36***	.23***	.18***	.23***	.23***	.23***	–

Note. COI = Child Opportunity Index; SES = Socioeconomic Status.

$^{***}p \leq .001$

$^{**}p \leq .01$

$^{*}p \leq .05$

$^{\dagger}p \leq .10$

As for the paths between SES indicators and SLEs (a_1), lower SLE scores were weakly predicted by higher household incomes, the presence of a partner, more educational neighbourhood resources, and lower poverty scores (reverse coded). The indirect effect of higher household income on lower *p* factor and specific attention factor scores via lower SLE scores was significant, and significantly (specific attention) or marginally (*p* factor) increased at higher levels of family obligation scores (see Table A2.9). Similarly, the indirect effect of the presence of a partner on higher *p* factor and specific internalizing factor scores via lower SLE scores was significant, and significantly (specific attention) or marginally (*p* factor) increased at higher levels of family obligation scores (see Table A2.9). The indirect effect of lower poverty scores on higher *p* factor and specific attention factor scores via lower SLE scores was significant and significantly increased at higher levels of family obligation scores (see Table A2.9). The indirect effects of neighbourhood social and economic resources on the *p* factor and specific externalizing scores, and caregiver education on specific attention scores, were not significant, since neighbourhood social and economic resources and caregiver education did not significantly predict SLEs at the first stage of mediation.

Table A2.9

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs) after controlling removing SLE items that overlapped with SES indicators

	Total Effect		Indirect Effect			IMM	
Path	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
Income							
P	-0.033	-0.051, -0.014	-0.015	-0.022, -0.008	45%	-0.001	-0.002, 0
Att	0.013	-0.001, 0.027	-0.003	-0.005, -0.002	23%	-0.001	-0.002, 0
Ext	-0.01	-0.023, 0.002	-0.003	-0.004, -0.001	30%	-0.006	-0.01, -0.001
Int	-0.006	-0.021, 0.009	-0.006	-0.008, -0.003	100%	0	0, 0.001
Education							
p	-0.015	-0.035, 0.004	-0.005	-0.012, 0.002	33%	0	-0.001, 0
Att	-0.021	-0.036, -0.005	-0.001	-0.002, 0	5%	0	-0.001, 0
Ext	-0.01	-0.023, 0.003	-0.001	-0.002, 0	10%	0	-0.001, 0
Int	0.001	-0.014, 0.017	-0.002	-0.004, 0.001	50%	0	0, 0.001
Single-partner							
p	-0.049	-0.099, 0	-0.107	-0.127, -0.089	46%	-0.006	-0.012, -0.001
Att	-0.027	-0.066, 0.01	-0.021	-0.027, -0.015	78%	-0.006	-0.011, -0.001
Ext	-0.025	-0.059, 0.007	-0.019	-0.024, -0.014	76%	-0.006	-0.01, -0.001
Int	0.021	-0.02, 0.061	-0.04	-0.049, -0.032	53%	0.002	-0.004, 0.007
COI Edu							
p	-0.4	-0.741, -0.063	-0.152	-0.269, -0.036	38%	-0.009	-0.025, -0.001
Att	-0.036	-0.303, 0.232	-0.03	-0.056, -0.007	83%	-0.009	-0.022, -0.001
Ext	-0.029	-0.255, 0.193	-0.027	-0.049, -0.006	93%	-0.008	-0.02, -0.001
Int	-0.021	-0.311, 0.255	-0.057	-0.102, -0.014	37%	0.002	-0.005, 0.013
COI Health							
p	-0.307	-0.93, 0.307	0.06	-0.157, 0.265	20%	0.004	-0.008, 0.022
Att	-0.265	-0.721, 0.215	0.012	-0.031, 0.053	5%	0.003	-0.008, 0.02

Ext	-0.214	-0.617, 0.206	0.01	-0.028, 0.047	5%	0.003	-0.007, 0.019
Int	-0.038	-0.523, 0.457	0.022	-0.058, 0.099	58%	-0.001	-0.015, 0.003
COI Social							
p	-0.093	-0.258, 0.063	0.034	-0.019, 0.089	37%	0.002	-0.001, 0.008
Att	0.027	-0.095, 0.149	0.007	-0.004, 0.018	26%	0.002	-0.001, 0.007
Ext	-0.119	-0.221, -0.015	0.006	-0.003, 0.016	5%	0.002	-0.001, 0.007
Int	-0.027	-0.153, 0.097	0.013	-0.007, 0.034	48%	-0.001	-0.005, 0.001
Poverty							
p	-0.056	-0.076, -0.035	-0.027	-0.036, -0.02	48%	-0.002	-0.003, 0
Att	-0.032	-0.048, -0.017	-0.005	-0.008, -0.004	16%	-0.002	-0.003, 0
Ext	-0.01	-0.024, 0.004	-0.005	-0.007, -0.003	50%	-0.001	-0.003, 0
Int	-0.011	-0.027, 0.004	-0.01	-0.014, -0.007	91%	0	-0.001, 0.002

Note. Att = specific attention factor; BS = Bootstrapped; COI = Child Opportunity Index; Ext = specific externalizing factor; IMM = Index of Moderated Mediation; Int = specific internalizing factor; P_M = percentage ratio of the indirect effect to the total effect.

2.5 Moderated Mediation with Correlated Psychopathology Factors

Figure A2.4 presents a moderated mediation model using internalizing, externalizing, and attention factor scores from the correlated factor model. Like in the bifactor version, higher SES scores weakly predicted lower SLE scores (a_1), and higher SLE scores moderately predicted higher internalizing, externalizing, and attention scores at equal strength (b_1). Higher SES scores weakly predicted lower internalizing, attention, and externalizing factor scores (c'), with predictions increasing modestly in strength in that order. Indirect effects (a_1b_1) of SES on externalizing and attention factors were negative and similar in strength to the indirect effect for the *p* factor in the bifactor version (Table A10). The indirect effect for the for internalizing scores did not reach significance, similar to the specific internalizing factor in the bifactor version.

Higher family obligation scores weakly predicted lower internalizing, externalizing, and attention factor scores with equal strength. The interaction between family obligation and SLE scores (b_2) was positive and equal strong for the externalizing and attention factors, but was not significant for the internalizing factor. Similarly, indices of moderated mediation (a_1b_3) were significant and negative for the externalizing and attention scores, but was not significant for internalizing scores (see Table A10).

Table A2.10

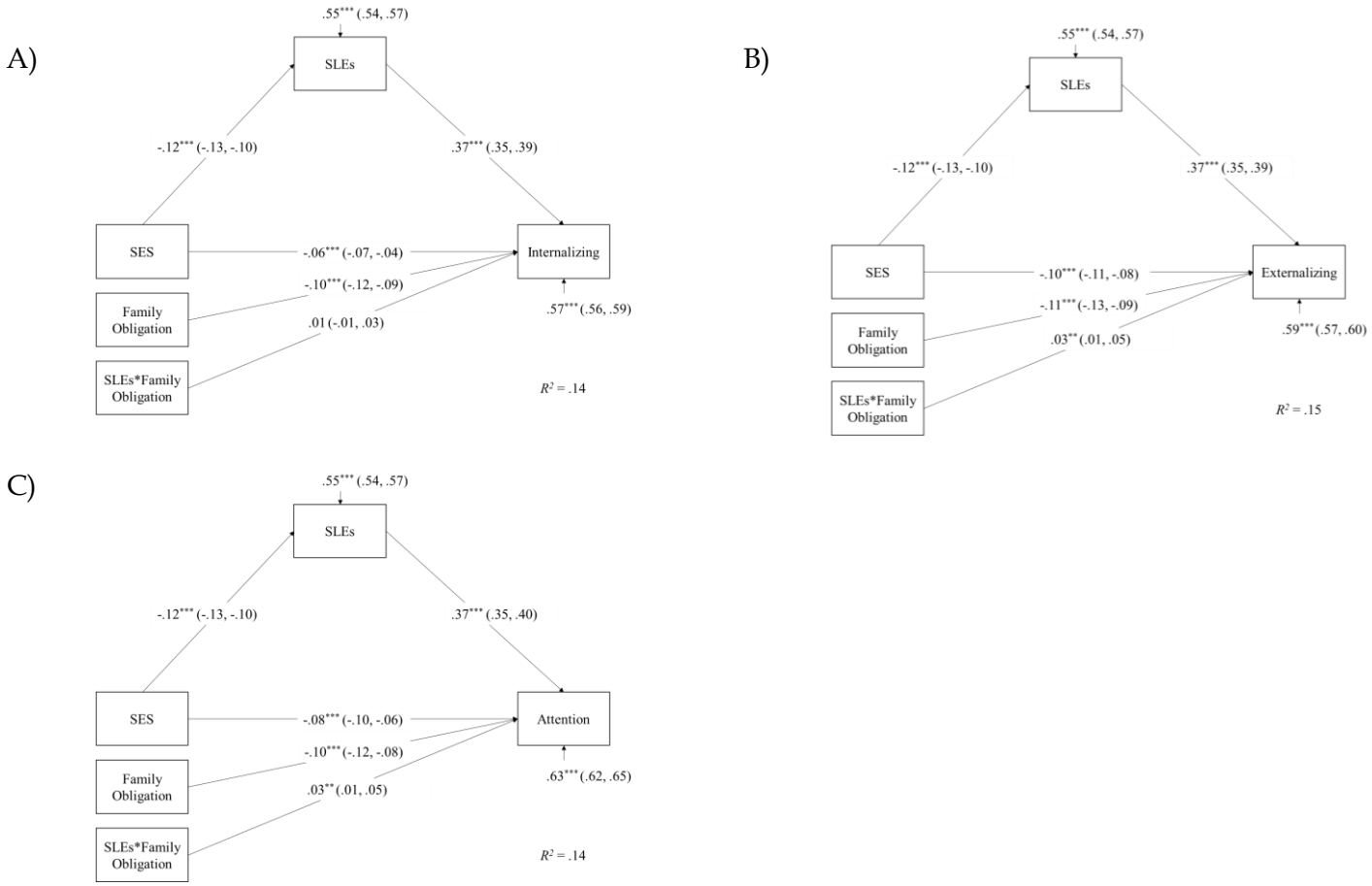
Standardized total effects, indirect effects, and the Index of Moderated Mediation for the correlated factor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs)

	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
Internalizing	-.10***	-.12, -.08	-.04***	-.05, -.04	43%	-.001**	-.004, .001
Externalizing	-.14***	-.16, -.12	-.04***	-.05, -.04	36%	-.003**	-.006, -.001
Attention	-.12***	-.14, -.10	-.04***	-.05, -.04	31%	-.003	-006, -.001

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.

Figure A2.4

Moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the internalizing factor (A), externalizing factor (B), and attention factor (C) from the correlated factors model



***$p \leq .001$; **$p \leq .01$; *$p \leq .05$; †$p \leq .10$

2.6 Cross-Informant and Parent-Only Models

A bifactor model with child-reported and parent-reported *p* factors and specific internalizing, externalizing, and attention factors showed healthy positive loadings (apart from the specific attention child factor; see Table A11) and acceptable absolute fit (but fell under acceptable incremental fit; see Table A12). Holding factor loadings constant between child and parent items worsened model fit (ΔAIC = -1557.09, ΔBIC = -1250.44, ΔBIC*n* = -1358.49; ΔCFI = -.01; ΔTLI = 0; ΔRMSEA = 0), violating assumptions of metric invariance. Holding both factor loadings and item response thresholds constant also worsened model fit (ΔAIC = -8269.81, ΔBIC = -7927.02, ΔBIC*n* = -8047.85; ΔCFI = .004; ΔTLI = -.04; ΔRMSEA = -.02), violating assumptions of scalar invariance.

A trifactor model (Wade et al., 2021) with a cross-informant *p* factor and cross-informant specific internalizing, externalizing, and attention factors; a child-reported *p* factor and specific internalizing, externalizing, and attention factors; and a parent-reported *p* factor and specific internalizing, externalizing, and attention factors, showed an excellent fit to the data (see Table A12), but imbalanced factor loadings (see Table A13). Specifically, parent items loaded more strongly onto the cross-informant *p* factor and specific attention factor than child items, whereas child items loaded more strongly onto the cross-informant specific externalizing factor than child items. Cross-informant specific internalizing factor loadings were similar among child and parent items. The parent-reported *p* factor and specific factors aside from externalizing showed variable and weak loadings. The child-reported *p* factor showed moderate and stable loadings, but child-reported specific factor loadings were variable and weak. Overall, these

results demonstrate that child- and parent-reported BPM items did not show strong enough overlap to be estimated by cross-informant factors (see also Watts et al., 2021).

Child and parent reported SLE items did not show metric invariance (ΔAIC = -254.45, ΔBIC = -135.69, ΔBIC*n* = -183.34; ΔCFI = .01; ΔTLI = .03; ΔRMSEA = -.002) or scalar invariance (ΔAIC = -9141.66, ΔBIC = -9014.93, ΔBIC*n* = -9065.77; ΔCFI = -.20; ΔTLI = -.16; RMSEA = .008; see Table A12 and Table A14 for factor loadings from the configural model). Furthermore, a bifactor model did not support a cross-informant SLE factor, but instead supported two specific factors related to each informant (see Table A14).

Child and parent reported MACVS items did not show metric invariance (ΔAIC = -587.72, ΔBIC = -468.92, ΔBIC*n* = -516.59; ΔCFI = 0; ΔTLI = 0; ΔRMSEA = 0) or scalar invariance (ΔAIC = -8441.99, ΔBIC = -7935.09, ΔBIC*n* = -8138.47; ΔCFI = -.02; ΔTLI = -.02; ΔRMSEA = .007; see Table A12 and Table A15 for factor loadings from the configural model). Furthermore, a bifactor model did not support a cross-informant family obligation factor, but a parent-dominated general factor and strong child-reported specific factor (see Table A15). Overall, there was no clear evidence for cross-informant SLE or family obligation factors, so I analyzed child- and parent-reported items separately.

I also replicated the moderated mediation model with parent-reported items only to check for consistency with the child-reported version. Figure A2.5 shows a moderated mediation model using parent-reported SES, SLE, family obligation, and bifactor psychopathology factor scores. Results were generally consistent with the

child-reported model. However, SES positively but weakly predicted specific internalizing and attention scores, despite not significantly predicting specific internalizing scores and marginally predicted specific attention scores in the child-reported model). Moreover, family obligation scores positively predicted specific externalizing scores, despite negatively predicting externalizing scores in the child-reported model. All indirect effects remained significant and were moderated by family obligation, except for the indirect effect on specific internalizing scores which was not moderated by family obligation in either parent- or child-reported models (Table A16).

Table A2.11

Standardized factor loadings for the bifactor model of the Brief Problem Monitor child-reported and parent-reported items

	Child				Parent			
Item	*p*	Ext	Int	Att	*p*	Ext	Int	Att
Acts young	0.55	-0.01			0.55	0.27		
Argues	0.59		0.47		0.72		0.45	
Can't finish things	0.60	0.16			0.70	0.37		
Can't concentrate	0.69	0.57			0.65	0.74		
Can't sit still	0.58	0.42			0.66	0.46		
Destroys things	0.56		0.32		0.71		0.28	
Disobedient (home)	0.56		0.38		0.75		0.45	
Disobedient (school)	0.56		0.30		0.69		0.26	
Worthless	0.60			0.57	0.61			0.51
Impulsive	0.73	-0.02			0.80	0.21		
Fearful	0.54	0.46		0.61	0.46			0.69
Guilty	0.55			0.52	0.38			0.66
Embarrassed	0.52			0.46	0.47			0.48
Distracted	0.71				0.70	0.58		
Stubborn	0.54		0.27		0.73		0.30	
Temper	0.60		0.44		0.71		0.39	
Threatens	0.56		0.48		0.73		0.31	
Unhappy	0.54			0.54	0.61			0.46
Worries	0.51			0.59	0.40			0.74
M	0.58	0.26	0.38	0.55	0.63	0.44	0.35	0.59

SD	0.06	0.25	0.08	0.05	0.12	0.20	0.08	0.12
ECV	0.64	0.07	0.10	0.18	0.64	0.11	0.07	0.18
ω	0.94	0.86	0.86	0.91	0.96	0.92	0.93	0.90
ω_H	0.81	0.13	0.27	0.45	0.82	0.27	0.18	0.53
Rel. ω	0.86	0.15	0.31	0.50	0.85	0.29	0.19	0.59
H	0.91	0.50	0.56	0.73	0.94	0.69	0.51	0.79
FD	0.93	0.84	0.75	0.89	0.93	0.91	0.70	0.90

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; ω = Omega; ω_H = Omega Hierarchical.

Table A2.12

Model fit statistics for measurement invariance testing and cross-informant (child and parent) factor models of the Brief Problem Monitor (BPM), Stressful Life Events (SLE) Checklist, and Mexican-American Cultural Values Scale (MACVS) items

Measure/Model	χ^2 (*df*)	RMSEA	CFI	TLI	AIC	BIC	BIC*n*
BPM							
Bifactor - configural	2592.76 (627)	.063 (.062-.064)	.87	.85	408,034.83	409,405.74	408,922.68
Bifactor - metric	27021.16 (661)	.063 (.062-.063)	.86	.85	409,591.92	410,656.18	410,281.17
Bifactor - scalar	32569.41 (699)	.067 (.066-.068)	.83	.83	417,861.73	418,583.26	418,329.02
Trifactor with cross-informant factors	3672.75 (551)	.024 (.023-.024)	.98	.98	404,071.63	405,719.48	404,994.93
SLE							
Bifactor - configural	20478.01 (464)	.065 (.064-.066)	.41	.37	215,332.86	215,839.78	215,636.39
Bifactor - metric	20073.08 (479)	.063 (.063-.064)	.42	.4	215,587.33	215,975.45	215,819.73
Bifactor - scalar	27179.59 (495)	.073 (.072-.074)	.21	.21	224,728.99	224,99.37	224,885.50
MACVS							
Bifactor - configural	14484.83 (464)	.055 (.054-.055)	.95	.95	702,419.81	703,687.07	703,178.60
Bifactor - metric	15114.35 (479)	.055 (.054-.056)	.95	.95	703,007.54	704,155.99	703,695.18
Bifactor – scalar	21663.02 (543)	.062 (.061-.063)	.93	.93	711,449.52	712,091.07	711,833.66

Note. AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; BIC*n* = sample-sized corrected Bayesian Information Criteria; CFI = Comparative Fit Index; *df* = Degrees of Freedom; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation; χ^2 = Chi-square.

Table A2.13

Standardized factor loadings for the trifactor model of the Brief Problem Monitor with cross-informant, child-reported, and parent-reported general and specific psychopathology factors

	Cross-Informant				Child				Parent			
Item	*p*	Ext	Int	Att	*p*	Ext	Int	Att	*p*	Ext	Int	Att
Child												
Acts young	0.17			0.17	0.52			-0.10				
Argues	0.31	0.41			0.52	-0.10						
Can't finish things	0.23			0.18	0.54			0.10				
Can't concentrate	0.28			0.32	0.63			0.48				
Can't sit still	0.24			0.27	0.52			0.33				
Destroys things	0.35	0.20			0.47	-0.17						
Disobedient (home)	0.39	0.27			0.44	-0.56						
Disobedient (school)	0.41	0.09			0.44	-0.34						
Worthless	0.24		0.50		0.55		0.55					
Impulsive	0.34			0.07	0.63			-0.04				
Fearful	0.13		0.63		0.54		-0.12					
Guilty	0.13		0.48		0.55		0.10					
Embarrassed	0.05		0.44		0.56		-0.01					
Distracted	0.25			0.26	0.66			0.40				
Stubborn	0.20	0.29			0.50	-0.04						
Temper	0.29	0.56			0.53	0.16						
Threatens	0.35	0.38			0.47	-0.12						
Unhappy	0.25		0.48		0.47		0.37					
Worries	0.08		0.58		0.54		-0.11					
Parent												
Acts young	0.53			0.27					0.15			0.00
Argues	0.69	0.17							0.19	0.44		
Can't finish things	0.67			0.36					0.21			-0.32

Can't concentrate	0.63			0.70					0.17			-0.10
Can't sit still	0.65			0.51					0.13			0.23
Destroys things	0.71	0.16							0.12	0.22		
Disobedient (home)	0.78	0.04							0.05	0.43		
Disobedient (school)	0.88	0.00							-0.23	0.08		
Worthless	0.50		0.43						0.52		-0.12	
Impulsive	0.78			0.22					0.19			0.15
Fearful	0.35		0.44						0.50		0.54	
Guilty	0.25		0.41						0.55		0.16	
Embarrassed	0.32		0.27						0.56		0.06	
Distracted	0.64			0.60					0.26			-0.08
Stubborn	0.56	0.16							0.46	0.41		
Temper	0.61	0.22							0.31	0.42		
Threatens	0.69	0.29							0.22	0.24		
Unhappy	0.50		0.43						0.47		-0.11	
Worries	0.25		0.40						0.59		0.32	
M	0.41	0.23	0.46	0.33	0.53	-0.17	0.13	0.20	0.29	0.32	0.14	-0.02
SD	0.22	0.15	0.09	0.19	0.06	0.23	0.27	0.24	0.22	0.14	0.26	0.20
ECV	0.34	0.04	0.11	0.07	0.22	0.02	0.02	0.02	0.10	0.03	0.02	0.01
ω	0.97	0.93	0.92	0.94	0.95	0.88	0.91	0.87	0.96	0.94	0.91	0.86
ω_H	0.55	0.15	0.51	0.24	0.67	0.05	0.03	0.07	0.18	0.16	0.04	0.00
Rel. ω	0.56	0.16	0.56	0.25	0.71	0.06	0.03	0.08	0.19	0.17	0.04	0.00
H	0.94	0.56	0.78	0.71	0.89	0.40	0.39	0.39	0.76	0.50	0.37	0.17
FD	0.96	0.81	0.90	0.94	0.93	0.76	0.85	0.75	0.93	0.85	0.78	0.89

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; ω = Omega; ω_H = Omega Hierarchical.

Table A2.149

Standardized factor loadings for the configural invariance and cross-informant bifactor models of the Stressful Life Events items reported by children and parents

	Configural		Cross-Informant		
Informant/Item	Child	Parent	General	S. Child	S. Parent
Child					
Family member had a drug/alcohol problem	0.55		0.37	0.40	
Family member had a mental/emotional problem	0.58		0.33	0.47	
Parents separated/divorced	0.41		0.74	0.09	
Family member died	0.24		0.10	0.21	
Family member was seriously injured	0.46		0.11	0.49	
Witnessed crime/accident	0.41		0.07	0.44	
Lost a close friend	0.45		0.11	0.46	
Close friend seriously sick/injured	0.45		0.03	0.53	
Negative change in parents' financial situation	0.61		0.48	0.38	
I got seriously sick	0.48		0.06	0.53	
I got seriously injured	0.45		0.05	0.52	
Parents argued more than before	0.55		0.43	0.34	
Parent lost their job	0.47		0.47	0.26	
Parent was away from home more	0.46		0.32	0.32	
Close friend died	0.48		0.06	0.53	
I was a victim of a crime/assault/violence	0.51		0.26	0.45	
Parent					
Family member had a drug/alcohol problem		0.54	0.51		0.19
Family member had a mental/emotional problem		0.53	0.43		0.29
Parents separated/divorced		0.57	0.99		-0.31
Family member died		0.13	0.11		0.08
Family member was seriously injured		0.29	0.17		0.28
Witnessed crime/accident		0.33	0.20		0.29

Lost a close friend		0.43	0.20		0.56
Close friend seriously sick/injured		0.24	0.02		0.45
Negative change in parents' financial situation		0.82	0.59		0.58
I got seriously sick		0.20	0.09		0.21
I got seriously injured		0.28	0.13		0.27
Parents argued more than before		0.59	0.51		0.19
Parent lost their job		0.71	0.52		0.52
Parent was away from home more		0.56	0.46		0.23
Close friend died		0.41	0.11		0.65
I was a victim of a crime/assault/violence		0.50	0.40		0.30
M	0.47	0.45	0.29	0.40	0.30
SD	0.08	0.19	0.23	0.13	0.23
ECV			0.47	0.30	0.23
ω	.82	.80	0.87	0.83	0.83
ω_H			0.51	0.60	0.36
Rel. ω			0.58	0.72	0.44
H	.83	.87	0.98	0.78	0.75
FD	.91	.93	0.99	0.89	0.93

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; S. Child = specific child-reported factor; S. Parent = specific parent-reported factor; ω = Omega; ω_H = Omega Hierarchical.

Table A2.15

Standardized factor loadings for the configural invariance and cross-informant bifactor models of the Mexican-American Cultural Values Scale completed by children and parents

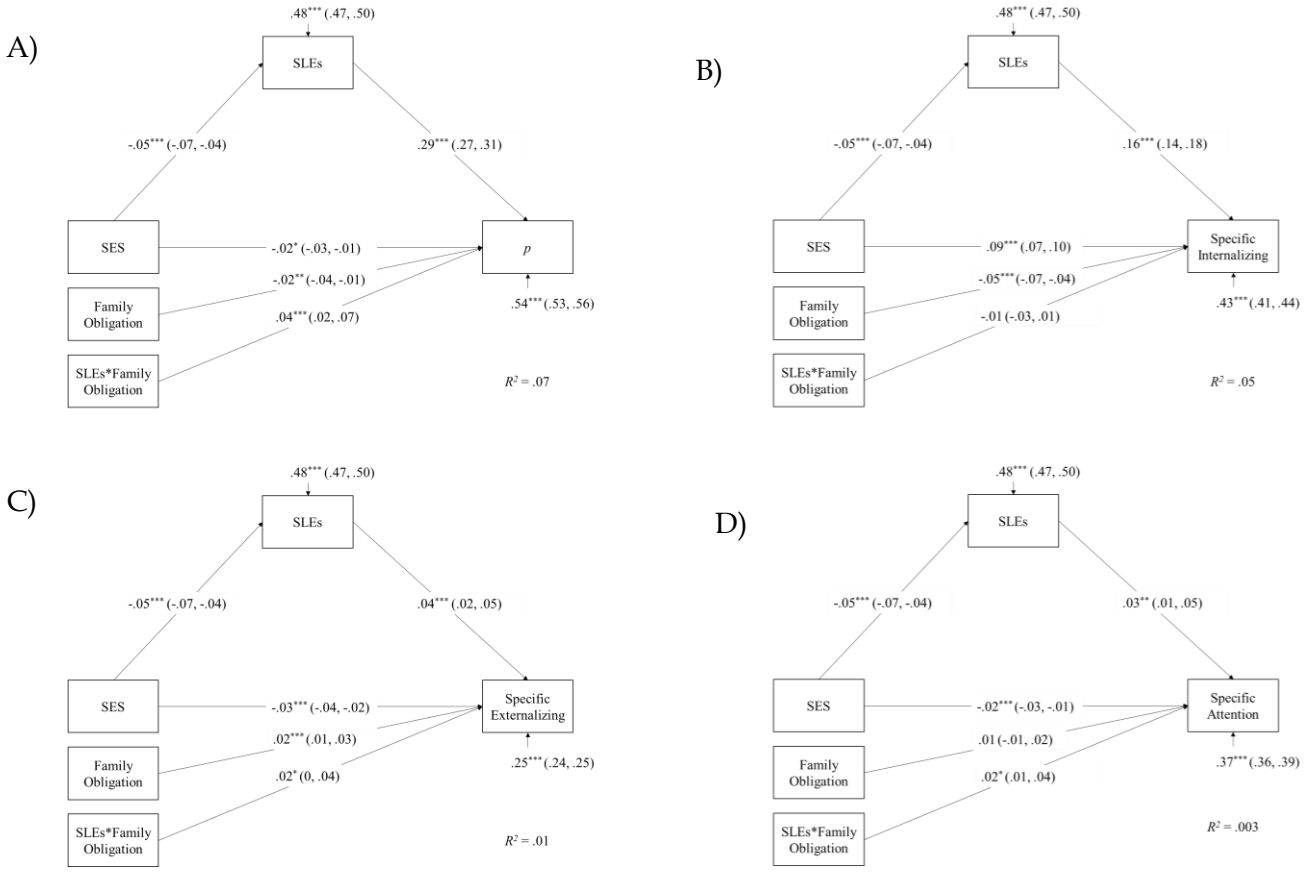
	Configural		Cross-Informant		
Informant/Item	Child	Parent	General	S. Child	S. Parent
Child					
Parents should teach children that family always comes first	0.66		0.65	0.63	
It is child's duty to care for parents when they get old	0.70		0.66	0.65	
Always do things to make parents happy	0.71		0.71	0.66	
Family provides a sense of security; will always be there	0.71		0.39	0.71	
Should help relative with financial hardship	0.71		0.28	0.70	
Family should ask for advice from relatives to make decisions	0.70		0.35	0.69	
Important to be united as a family	0.78		0.66	0.77	
Should share home with relatives if they need a place to stay	0.69		0.19	0.69	
Important to have close relationships with extended family	0.74		0.37	0.73	
Older kids should take care of younger brothers and sisters	0.71		0.57	0.69	
Always be good; represent the family.	0.79		0.88	0.75	
Celebrations are important because the family comes together	0.72		0.42	0.71	
Parents should make sacrifices so children have better life	0.65		0.52	0.62	
Always think about family when making important decisions.	0.78		0.54	0.76	
Important for family to show love and affection	0.75		0.39	0.74	
Work hard; reflects on the family	0.79		0.81	0.76	
Parent					
Parents should teach children that family always comes first		0.73	0.19		0.35
It is child's duty to care for parents when they get old		0.64	0.28		0.18
Always do things to make parents happy		0.62	0.26		0.07
Family provides a sense of security; will always be there		0.67	0.07		0.60
Should help relative with financial hardship		0.63	0.13		0.67
Family should ask for advice from relatives to make decisions		0.58	0.15		0.52

Important to be united as a family		0.78	0.12		0.44
Should share home with relatives if they need a place to stay		0.58	0.09		0.69
Important to have close relationships with extended family		0.68	0.15		0.64
Older kids should take care of younger brothers and sisters		0.69	0.17		0.40
Always be good; represent the family.		0.81	0.24		0.12
Celebrations are important because the family comes together		0.68	0.16		0.58
Parents should make sacrifices so children have better life		0.62	0.20		0.35
Always think about family when making important decisions.		0.68	0.16		0.42
Important for family to show love and affection		0.65	0.11		0.55
Work hard; reflects on the family		0.79	0.19		0.17
M	0.72	0.68	0.35	0.70	0.42
SD	0.04	0.07	0.23	0.05	0.20
ECV			0.32	0.47	0.20
ω	0.95	0.93	0.95	0.98	0.81
ω_H			0.39	0.63	0.70
Rel. ω			0.42	0.64	0.86
H	0.95	0.95	0.92	0.94	0.84
FD	0.97	0.97	0.60	0.97	0.92

Note. Att = Attention; ECV = Explained Common Variance; Ext = Externalizing Factor; *FD* = Factor Determinacy; *H* = Construct Reliability; Int = Internalizing Factor; Rel ω = Reliable Omega; S. Child = specific child-reported factor; S. Parent = specific parent-reported factor; ω = Omega; ω_H = Omega Hierarchical.

Figure A2.5

Moderated mediation path diagrams with standardized path coefficients and 95% bootstrapped confidence intervals (in parentheses) for the p factor (A), specific internalizing factor (B), specific externalizing factor (C), and specific attention factor (D) using parent-reported items only



***$p \leq .001$; **$p \leq .01$; *$p \leq .05$; †$p \leq .10$

Table A2.16

Standardized total effects, indirect effects, and the Index of Moderated Mediation for the bifactor dimensions predicted by Socioeconomic Status (SES) via Stressful Life Events (SLEs) using parent-reported indicators only

	Total Effect		Indirect Effect			IMM	
Factor	*B*	BS 95% CI	*B*	BS 95% CI	P_M	*B*	BS 95% CI
p	-0.03	-0.05, -0.02	-0.014	-0.02, -0.01	45%	-0.002	-0.004, -0.001
Attention	-0.02	-0.04, -0.01	-0.001	-0.003, -0.001	4%	-0.001	-0.002, 0
Externalizing	-0.03	-0.04, -0.02	-0.002	-0.003, -0.001	6%	-0.001	-0.002, 0
Internalizing	0.02	0.07, 0.09	-0.008	-0.01, -0.005	NA	0.001	0, 0.002

Note. BS = Bootstrapped; IMM = Index of Moderated Mediation; P_M = percentage ratio of the indirect effect to the total effect.