The Development of the General Factor of Psychopathology 'p Factor' Through Childhood and Adolescence

Aja Louise Murray[1] · Manuel Eisner[1] · Denis Ribeaud[2]

Abstract Recent studies have suggested that the structure of psychopathology may be usefully represented in terms of a general factor of psychopathology (p-factor) capturing variance common to a broad range of symptoms transcending diagnostic domains in addition to specific factors capturing variance common to smaller subsets of more closely related symptoms. Little is known about how the general co-morbidity captured by this p-factor develops and whether general co-morbidity increases or decreases over childhood and adolescence. We evaluated two competing hypotheses: 1) dynamic mutualism which predicts growth in general co-morbidity and associated p-factor strength over time and 2) p-differentiation which predicts that manifestations of liabilities towards psychopathology become increasingly specific over time. Data came from the Zurich Project on the Social Development of Children and Youths (z-proso), a longitudinal study of a normative sample (approx. 50 % male) measured at 8 time points from ages 7 to 15. We operationalised general co-morbidity as p-factor strength in a bi-factor model and used omega hierarchical to track how this changed over development. In contrast to the predictions of both dynamic mutualism and p-differentiation, p-factor strength remained relatively constant over the studied period suggesting that such processes do not govern the interplay between psychopathological symptoms during this phase of development. Future research should focus on earlier phases of development and on factors that maintain the consistency of symptom-general covariation across this period.

Keywords Co-morbidity · Development · General factor of psychopathology · p-factor · Dynamic mutualism

Recent challenges to traditional notions of psychopathological disorders as distinct categories of impairment have prompted a re-consideration of psychopathology taxonomies. These challenges have been based on high degrees of co-morbidity that not only transcend diagnostic boundaries but also seem to be general to almost all symptoms of common mental disorders. The mechanisms by which psychopathological disorders develop these inter-relations, however, remains poorly understood. In this study, we test competing hypotheses regarding the development of general co-morbidity from late childhood into adolescence as an initial foray into this question.

Psychopathologies that were classically conceptualised as unrelated have been shown to exhibit substantial co-morbidity (Krueger and Markon 2006) with, for example, almost half of individuals who meet diagnostic criteria for one disorder also meeting diagnostic criteria for another (Kessler et al. 2005). This co-variation cuts across not only specific diagnostic categories, but also higher-order dimensions of psychopathology. The inter-correlations between the broad internalising (e.g., depression and anxiety), externalising (e.g., substance use, delinquency aggression, and hyperactivity) and thought disorder (e.g., psychosis) dimensions have, for example, been estimated as > 0.5 (Lahey et al. 2004; Wright et al. 2013).

✉ Aja Louise Murray
    am2367@cam.ac.uk

[1]   Violence Research Centre, Institute of Criminology, University of Cambridge, Cambridge CB3 9DA, UK

[2]   Criminological Research Unit, Chair of Sociology, Swiss Federal Institute of Technology Zurich (ETH), Zurich, Switzerland

There have been attempts to develop updated nosologies capable of capturing this trans-diagnostic inter-relatedness. One has centred on the idea of a general factor of psychopathology, labelled the *p*-factor by Caspi et al. (2014). The *p*-factor provides a statistical summary of the variance common to psychopathological symptoms across disorders and diagnostic domains. In methodological terms, the *p*-factor approach involves fitting a bi-factor measurement model to psychopathology data (e.g., see Holzinger and Swineford 1937; Murray and Johnson 2013). The basic form of a bi-factor model is shown in Fig. 1. Each item loads on two factors: one general factor that is common to most or all items and one specific factor common to a subset items that are related to one another over and above their relation captured by the general factor. The general factor is orthogonal to the specific factors and - by convention but not necessity - the specific factors are usually also mutually orthogonal. This specification allows the common variance amongst a set of items to be partitioned into that which is common to all items and that which is common to more specific domains.

When applied to psychopathology data, the general factor is the *p*-factor and the specific factors are most commonly internalising and externalising with other specific factors (such as thought disorder) included depending on the extent to which these domains are represented by available items. A bi-factor psychopathology model may then be compared in terms of its fit to alternative structural models, interpreted in terms of the relative magnitudes of *p*-factor and specific factor loadings, or used to obtain estimates of the unique (i.e., controlling for the other factors in the model) relations between the *p*-factor or specific factors and external variables.

Studies based on the bi-factor approach have yielded various insights into the associations among psychopathology symptoms and related external variables. Bi-factor
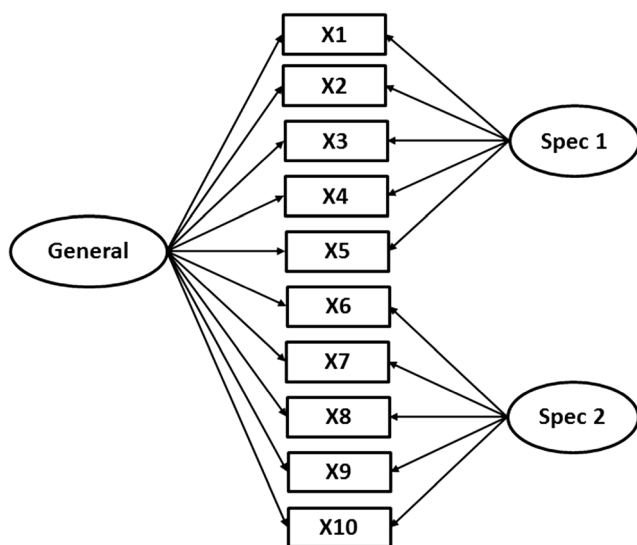

Fig. 1 Example bi-factor model

measurement models have tended to yield a very good fit to psychopathology data by conventional model fit criteria, leading to the conclusion that they may provide useful descriptions of the latent structure of mental disorders (Caspi et al. 2014; Laceulle et al. 2015; Lahey et al. 2012, 2015; Tackett et al. 2013). Examining parameter estimates from the model, it can be seen that not only some but much of the variance in psychopathological symptoms is attributable to the *p*-factor (i.e., shared among symptoms traditionally assumed to be manifestations of distinct domains). The *p*-factor does not, however, completely account for symptom covariation. In the case of internalising and externalising factors at least, sufficient common variance after extracting the *p*-factor usually remains for these factors to be maintained, albeit with attenuated loadings.

Building on these findings, a key goal has been providing an interpretation of the covariance that the *p*-factor captures. The most substantive interpretation and the one that has the potentially greatest impact on how psychopathologies are (re-)conceptualised is that it represents the effects of shared etiological factors such as pleiotropic genetic effects, personality traits, or broad-acting environmental exposures (e.g., Lahey et al. 2011; Stochl et al. 2015). Indeed, as Patalay et al. (2015) noted, sets of risk factors tend to be quite similar across different disorders, although few studies have addressed this question within the *p*-factor framework. In one study, Tackett et al. (2013) found substantial phenotypic and genetic overlap between the *p*-factor and negative emotionality (but not other dispositional traits), which they suggested may make it a candidate shared etiological factor underlying the *p*-factor. However, what these analyses could not rule out and what still remains to be addressed is whether negative emotionality is better considered a common outcome of psychopathological disorders rather than a common cause. That is, the correlation between the *p*-factor and negative emotionality could reflect the fact that distress occurs as a result of almost any psychopathological symptom.

An alternative perspective is that rather than reflecting a common cause or set of common causes, the *p*-factor is the emergent result of a network of symptoms that interact locally with one another (see Borsboom and Cramer 2013). Borsboom et al. (2011) provide examples of plausible causal chains of symptoms that run across different disorders such as sleep deprivation (depression symptom) impacting on a series of mediating symptoms that ultimately give rise to irritability (generalised anxiety disorder symptom). This perspective also allows for external influences such as adverse life events initiating these causal chains; however, the key point is that influence spreads through local specific interactions rather than by simultaneously affecting a broad range of symptoms at once (i.e., acting as a common cause). Such local interactions among symptoms could produce data consistent with a *p*-factor, therefore, the ability to fit a model including a *p*-factor does not imply that it represents the underlying

cause(s) of symptom covariation (Borsboom and Cramer 2013; van Der Maas et al. 2006).

A remaining question and one that has the potential to shed light on the etiological roots of general co-morbidity is how the widespread covariance between symptoms develops over time. Within the bi-factor approach, it is possible to assess whether general co-morbidity, as measured by the strength of the p-factor is constant from early in life, whether it grows over time as symptoms become increasingly correlated, or whether it decreases as symptoms become increasingly differentiated.

The idea that general co-morbidity grows in strength over time is consistent with a dynamic mutualism process whereby symptoms both across and within domains can reinforce one another through local interactions. There is no set of common causes providing a shared etiology for different symptoms but a network of contingencies and direct causal interactions between symptoms (see Borsboom et al. 2011 for examples). Over time, these local interactions can lead to a growth in symptom inter-correlations such that symptoms that were initially minimally correlated can end up substantially correlated (van Der Maas et al. 2006). A dynamic mutualism explanation has recently been posited as an explanation for the p-factor (Caspi et al. 2014). In general terms a dynamic mutualism model can be characterised as comprising two parts: a dynamic part and an interaction part. The dynamic part describes the development of symptoms over time and the interaction part describes the causal linkages between symptoms. Different patterns of symptom interactions would be expected to produce variations in patterns of inter-correlations and factor structures. For example, if the interactions between symptoms within the internalising and externalising domains are even slightly stronger than those between these domains, then one would expect these to emerge as broad dimensions in addition to general co-morbidity. Here, increasing co-morbidity could occur both within and between domains over time, leading to a strengthening of both p- and specific factors across development.

The idea of a dynamic mutualism process underlying general co-morbidity development is consistent with much of the developmental literature which has shown that relations between symptoms in different domains can emerge over time without necessarily having a common cause. These kinds of hypotheses have been discussed under the banner of 'cascade models' in which symptoms in one domain spread to another over time (Masten and Cicchetti 2010). As an example, externalising and internalising may be relatively independent initially; however, over time externalising behaviours can promote negative social experiences and impair academic performance leading, in turn, to internalising symptoms (van Lier et al. 2012). As these kinds of processes play out during the course of development, symptoms across multiple domains of psychopathology could become increasingly correlated.

While they might begin very early in life, they could continue to influence psychopathology symptom inter-relations into later childhood and early adulthood.

The direct alternative possibility for the course of general co-morbidity over development is that psychopathological symptoms become increasingly differentiated over time. Here, if the p-factor represents a liability to any kind of psychopathology, then, as individuals develop, the manifestation of that liability could become increasingly specific (e.g., Patalay et al. 2015). This could be due to an increasing strength of specific factors of psychopathology at the expense p-factor strength or it could be due to increasing differentiation at the symptom level. In the former case, a general liability would be increasingly replaced by a liability for symptoms within specific dimensions such as externalising and internalising. In the latter case a general liability would become manifested in increasingly idiosyncratic symptom patterns which may but need not also be accompanied by a decline in specific factor strength.

There is some evidence for symptom differentiation over development although whether it is attributable to declining influences of general or specific factors has not been tested. For example, based on factor analytic evidence, Cole et al. (1998) found that internalising symptoms were best characterised as a single dimension for children in the third grade, but as two dimensions: anxiety and depression, when the same children were in the sixth grade. Similarly, Lahey et al. (2004) found that items measuring oppositional defiant disorder and hyper-activity-impulsivity tended to load on the same factor in younger children but on separate factors in older children. This kind of differentiation would also be possible within a system characterised by dynamic mutualism, however, it would generally be expected to occur later in development, after an initial period of growth in p-factor strength (e.g., see van Der Maas et al. 2006).

These possibilities regarding the development of general co-morbidity can be compared by examining the relative amount of variance in psychopathology symptoms accounted for by a p-factor over the course of development. It was, therefore, the aim of the current study to use developmental data from individuals measured in childhood through to adolescence to establish which of the possibilities above provides the best account of the development of the general co-morbidity.

Method

Participants and Measures

Data came from the Zurich Project on the Social Development of Children and Youths (z-proso), a longitudinal cohort and intervention study focussed on the antecedents of violence and

aggression. The target sample was 1675 children from 56 public primary schools. Schools were selected according to a stratified random sampling procedure that took into account school location and size. All children who entered first grade in 2004 in one of these schools were invited to participate via their parents. Informed consent was obtained from parents at the beginning of data collection and from the children from age 13 onwards. Ethical approval for the study was obtained from the relevant institutional ethics bodies.

Overall 1572 of the target sample participated. Approximately half of the initial sample was male and the median date of birth was 22/10/1997. The sample is ethnically diverse: approximately 10 % of the children were born in Switzerland but only 54 % had parents that were both born in Switzerland. Only 38.4 % of the children were of Swiss nationality, after which the most commonly reported nationalities were: Italy (8.8 %), Serbia-Montenegro (8.7 %), Germany (6 %) and Portugal (5 %).

In terms of socioeconomic status, at baseline 76.7 % of the youths' male primary caregivers for which data were available were in full-time employment (8.8 % unemployed). The highest educational levels of male primary caregivers were: 21 % mandatory school or less, 35.2 % apprenticeship, 7.8 % A-levels, 15.5 % higher vocational education, and 16 % University. In terms of household finances, 17.7 % reported experiencing financial difficulties in the last year. Participation was not completely random and in general can be characterised as having resulted in an under-representation (with respect to the target sample) of individuals whose parents did not speak German as a first language.

Teacher ratings were obtained at eight time points covering the entire age range of compulsory schooling in Zurich. The median ages of the children at these time points were: 7.45, 8.23, 9.21, 10.70, 11.60, 12.63, 13.88, and 15.68. We henceforth label these measurement waves based on rounding down to the nearest whole age year. The intervention components took place early in the study when the children were in grades 1 to 3 and involved separate child and parent programmes. However, because intervention effects were not supported (see Averdijk et al. 2016; Malti et al. 2011) we treated the interventions as part of the natural milieu of the children.

Measures

To rate the target youth's behaviour, teachers completed the Social Behaviour Questionnaire (SBQ). The SBQ is composed of around 45 items, depending on the measurement wave because new items were added and others removed to maintain developmental appropriateness. We focussed on the 39 of these 45 items that were completed by teachers across all eight waves to allow unambiguous comparisons across waves. These items measured the concepts of prosocial behaviour

comprising helping and empathy; internalising behaviour comprising anxiety and depression; attention-deficit hyperactivity disorder (ADHD) comprising AD and HD; non-aggressive externalising behaviour comprising stealing, lying, vandalism and opposition/defiance; and aggression comprising physical aggression, indirect aggression, instrumental aggression/dominance and reactive aggression. All were measured on a five point Likert scale from *Never* to *Very often*.

The scale was first used by Tremblay et al. (1991) and was an amalgamation of two pre-existing scales: 28 items from the Preschool Behavior Questionnaire (Behar and Stringfield 1974), itself an adaptation of the Children's Behavior Questionnaire (Rutter 1967) and 10 items from the Prosocial Behavior Questionnaire (Weir and Duveen 1981). The version used in z-proso differs from this scale in that additional items have been added to facilitate the measurement of several sub-types of aggression. In addition, the scale was administered in German. Previous psychometric analyses have supported the ability of this version to reliably measure psychopathology from moderately low to very high trait levels, consistent with a dimensional approach to psychopathology measurement (Murray et al. 2016). For all analyses the items were (re-) coded in the direction of higher item scores indicating higher levels of psychopathology. Thus, scoring of the pro-sociality items was reversed.

Statistical Procedure

Items were treated as continuous which is a reasonable strategy for five-point scales provided that the response distributions are broadly symmetrical (Rhemtulla et al. 2012). Nonetheless, as an additional check we also estimated a subset of models from the main analysis using polychoric correlations and results were only trivially different.

As it was important to ensure that any changes over time were not due to differential attrition, we used Bayesian multivariate imputation to deal with missingness, employing the mice package in R statistical software (Van Buuren and Groothuis-Oudshoorn 2011; R Core Team 2014). Details of this procedure are provided in Supplementary Materials 1.

Within-Group *p*-Factor Strength

To assess whether the strength in the *p*-factor changed over time, we first estimated higher-order exploratory factor analyses at each time point. The number of specific factors included in these was guided by the preliminary analyses outlined in Supplementary Materials 2. To estimate *p*-factor strength at each time point, we used the *p* and specific factor loadings from a Schmid-Leiman transformation (Schmid and Leiman

1957; Yung et al. 1999) to compute an index of p-factor saturation or 'omega hierarchical' (ω_h; McDonald 1999). ω_h is computed as:

$$\omega_h = \frac{\left(\sum \lambda_{iP}\right)^2}{\left(\sum \lambda_{iP}\right)^2 + \left(\sum \lambda_{iS1}\right)^2 + \left(\sum \lambda_{iS2}\right)^2 + ... \left(\sum \lambda_{iSK}\right)^2 + \sum \theta_i^2}, \quad (1)$$

where λ_{iP} denotes the p-factor loading of item i; λ_{iS1} to λ_{iSK} denote the specific factor loadings of item i for specific factors 1 to K; and θ_i^2 denotes the error variance from item i. ω_h thus provides an index of the proportion of total (or summed) score variance that is attributable to the p-factor. The numerator is the variance due to the p-factor and the denominator is the variance of the summed score for all items. ω_h can be thought of as a measure of the strength of p-factor controlling for the specific factors (Reise et al. 2013). This approach to estimating changes in the strength of a general factor over time has previously been used to evaluate the dynamic mutualism hypothesis in cognitive ability research (Gignac 2014). ω_h can be computed in an analogous manner for the specific factors in the model in order to obtain an estimate of the amount of variance in the sum score of all items that is attributable to a given specific factor. These are calculated by replacing the numerator of eq. 1 with the square of summed loadings for the relevant specific factor.

Computing ω_h from the results of a Schmid-Leiman transformation of a higher-order model is not identical to computing it from a direct bi-factor model like those estimated in previous p-factor studies because unlike the direct bi-factor model, it maintains the proportionality constraints of a higher-order model. Specifically, the higher-order model is nested within the bi-factor model with the higher-order model equivalent to a bi-factor model in which the ratios of the item variance attributable to the p-factor and the relevant specific factor constrained equal across items belonging to the same specific factor (Yung et al. 1999). Revelle and Wilt (2013) reviewed various procedures that have been proposed for assessing the strength of a general factor and recommended the Schmid-Leiman transformation of a higher-order model approach as the most appropriate and more appropriate than using a direct confirmatory bi-factor model. Their rationale was that a CFA approach fitting a direct bi-factor model (rather than an indirect model using a Schmid-Leiman transformation) is more liable to over-estimate-general factor strength and, in turn ω_h, especially if the data are not simple structured (Revelle and Wilt 2012, 2013). This is likely due to the fact that CFA models constrain many loadings to zero and this unmodelled covariance may inflate p-factor loadings but to a greater extent in a CFA versus EFA model and in a bi-factor versus Schmid-Leiman transformed higher-order model (e.g., Murray and Johnson 2013).

There were also reasons to prefer an exploratory over a confirmatory approach in the current study. First, there is little previous research to guide the appropriate specification of a CFA model for the SBQ items, therefore, it is important to ensure that any changes in p-factor strength are not only associated with specific modelling constraints. Second, an exploratory approach allows that the factor model for the SBQ items to vary quite freely across measurement waves while still allowing for an estimation of p-factor strength. Given the array of developmental changes that occur between the ages of 7 and 15, this approach would appear more defensible than attempting to fit a similar or identical factor structure across all eight time points.

p-Factor Stability

Finally, we computed the stability of factor scores estimated from the same factor models used to compute the ω_h values. Factor scores were estimated using the method described in ten Berge et al. (1999). The adequacy of factor scores was evaluated using the correlation between scores and latent factors criterion (Grice 2001). The stabilities of the factor scores were corrected for attenuation due to unreliability based on these correlation.

Results

Within-Group Structure

p –Factor and Specific Factor Strength Change

Average sample sizes and item ICCs are provided in Table 1. We used a model developed from preliminary EFA analyses of the wave Age 10 (and replicated in waves Age 7 and Age 15) data to explore the development of the p-factor over time (see Supplementary Materials 2). Based on these, we extracted four specific factors and then one p-factor at every time point. The oblique factor correlation matrices from which the p-factors were extracted are provided in Supplementary Materials 3. ADHD and Aggression were consistently more strongly correlated than the other factors (between 0.41 and 0.50) which correlated with one another between 0.11 and 0.32. The Schmid-Leiman factor solutions at the eight measurement points are provided in Tables 2 and 3.

There was some fluctuation in the content of the p-factor over time but no obvious trend in any direction. Generally, none of the items had consistently strong relations to the p-factor but some of the strongest were from the ADHD domain and, to a lesser extent, the aggression domain. The internalising and pro-sociality items tended not to have strong relations with the p-factor and were instead more strongly related to the relevant specific factor.

Table 1 *p*-factor and specific factor strengths across waves

p-factor and specific factor strength

Wave	Average N (SD)	Average item ICC	ω_h *p*-factor	ω_h Aggression	ω_h ADHD	ω_h Pro-sociality	ω_h Internalising
Age 7	1333.31	0.13	0.59	0.17	0.07	0.06	0.06
Age 8	1318.05	0.20	0.59	0.18	0.06	0.06	0.06
Age 9	1289.85	0.18	0.60	0.18	0.06	0.06	0.06
Age 10	1261.23	0.13	0.64	0.17	0.05	0.05	0.05
Age 11	1058.7	0.10	0.59	0.19	0.06	0.05	0.06
Age 12	973.92	0.07	0.61	0.15	0.07	0.05	0.06
Age 13	1244.38	0.15	0.56	0.18	0.07	0.07	0.06
Age 15	1271.10	0.17	0.53	0.21	0.06	0.07	0.06

Note. Average N is the mean sample size across the 39 SBQ items

ω_h is the proportion of total score variance attributable to the relevant factor (general or specific)

The magnitudes of ω_h for the *p*-factor and each specific factor at each measurement wave are provided in Table 1 and plotted in Fig. 2. The ω_h values showed a slight peak at Age 10, then a gentle decline but stayed within a relatively narrow range of values. The ω_h values for each the specific factors also showed little variation over time. Overall, these results suggest that patterns of co-morbidity remain quite consistent between the ages of 7 and 15.

Estimating *p*-factor scores from the above-described factor models, the correlation between factor scores and latent factors fell in the range 0.78 -0.79, except for at Age 10 where this value was slightly higher at 0.82. These values fall short of the minimum recommended value of 0.90 (Gorsuch 1983). The standardised autoregressive coefficients for *p*-factor scores ranged from 0.10 (Age 13 regressed on Age 12) up to 0.33 (Age 9 regressed on Age 8) suggesting little stability in *p*-factor scores across time. Stability is, however, limited by the correlations between latent factors and factor scores and correcting for this unreliability, the autoregressive coefficients were, from the first to last measurement wave: 0.40, 0.43, 0.23, 0.26, 0.39, 0.12 and 0.18. The corresponding attenuation-corrected autoregressive coefficients for the Aggression factor were: 0.56, 0.54, 0.35, 0.42, 0.52, 0.20 and 0.25; for ADHD they were: 0.58, 0.57, 0.34, 0.52, 0.61, 0.23 and 0.35; for Pro-sociality they were: 0.47, 0.50, 0.23, 0.31, 0.43, 0.07 and 0.23; and for Internalising they were: 0.39, 0.42, 0.18, 0.37, 0.39, 0.20 and 0.30.

Discussion

In this study, we extended previous findings that the latent structure of psychopathology data can be characterised as involving co-morbidity that is both general and domain-specific. Based on factor analytic evidence, we judged that a bi-factor model with a general *p*-factor, together with the specific factors of internalising, aggression, ADHD and pro-social behaviour provided a good representation of the structure of psychopathology in a normative sample of individuals measured at eight time points between the ages of 7 to 15. The relative strength of the *p*-factor and specific factors varied within a relatively narrow range over this time period and did not show an overall systematic increase or decrease with time. Such trajectories are not consistent with simple versions of a dynamic mutualism process of *p*-factor growth over time which would predict an increasingly strong *p*-factor with time. Nor are they consistent with *p*-differentiation: a process of increasing specificity in the expression of a general liability for psychopathology. Rather, our results suggest that from the point at which children enter school until adolescence, the extent to which a diversity of psychopathological behaviours within and between domains are co-morbid remains quite constant.

The fact that a bi-factor model with both a general *p*-factor and several specific factors provided a good representation of psychopathology data provides a conceptual replication of several previous studies (Caspi et al. 2014; Lahey et al. 2011, 2012; Laceulle et al. 2015; Stochl et al. 2015; Tackett et al. 2013). Though the content of the specific factors will vary across studies, the finding that a general bi-factor structure describes psychopathology data well appears to be robust. The fact that these results hold across the different sets of symptoms included in independent studies supports the generality of the *p*-factor. For example, although most studies have included 'internalising' and 'externalising' specific factors, Caspi et al. (2014) and Laceulle et al. (2015) added a 'thought disorder' factor, Stochl et al. (2015) specified a psychotic experiences factor but no externalising factor, and Lahey et al. (2012) split the internalising factor into 'distress' and 'fear' factors. The current study allows further generalisation by including an extensive set of aggression and pro-sociality items. Although statistical criteria have supported the

Table 2 Schmid-Leiman loadings for *p*-factor over time

Item	Abbreviated content	Measurement wave							
		Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
SBQ02	Nervous, tense	0.39	0.37	0.42	0.44	0.38	0.35	0.33	0.34
SBQ03	Fearful/anxious								
SBQ04	Worried						0.30		
SBQ05	Unhappy/sad/depressed	0.32		0.31		0.32	0.33		
SBQ06	Not so happy	0.37		0.33	0.30	0.32	0.36		
SBQ07	Anhedonic						0.30	0.33	
SBQ08	Miserable/ distressed/unhappy		0.30			0.33	0.30		
SBQ10	Impulsive	0.55	0.54	0.56	0.63	0.52	0.51	0.56	0.54
SBQ11	Impatience with turn-taking	0.52	0.51	0.51	0.59	0.49	0.49	0.52	0.52
SBQ12	Restless/hyperactive	0.53	0.53	0.53	0.69	0.51	0.48	0.54	0.56
SBQ13	Fidgets	0.50	0.51	0.52	0.66	0.49	0.45	0.50	0.52
SBQ14	Can't settle	0.52	0.53	0.54	0.77	0.59	0.50	0.58	0.59
SBQ15	Distractible	0.55	0.54	0.54	0.79	0.59	0.49	0.59	0.61
SBQ16	Can't sustain concentration	0.56	0.52	0.53	0.80	0.60	0.50	0.60	0.60
SBQ17	Inattentive	0.54	0.51	0.53	0.74	0.56	0.49	0.56	0.56
SBQ26	Destroys own things	0.38	0.35	0.41	0.35	0.36	0.37	0.36	0.38
SBQ27	Disobedient	0.52	0.48	0.49	0.50	0.51	0.51	0.50	0.53
SBQ30	Ignores you	0.50	0.45	0.46	0.49	0.46	0.49	0.44	0.45
SBQ31	Destroys others' things	0.35	0.35	0.42	0.37	0.35	0.42	0.37	0.41
SBQ32	Lies and cheats	0.42	0.42	0.45	0.44	0.38	0.42	0.42	0.44
SBQ33	Gets into fights	0.54	0.54	0.54	0.48	0.49	0.48	0.46	0.38
SBQ34	Physically attacks	0.56	0.56	0.54	0.49	0.50	0.51	0.48	0.40
SBQ35	Kicks, bites, hits	0.54	0.54	0.50	0.43	0.44	0.44	0.44	0.33
SBQ36	Cruel, bullies	0.47	0.47	0.48	0.42	0.45	0.50	0.46	0.41
SBQ37	Threatens	0.47	0.49	0.50	0.43	0.45	0.54	0.45	0.40
SBQ41	Volunteers to help	0.37		0.30	0.33		0.31	0.32	
SBQ42	Tries to stop disputes	0.40	0.32	0.36	0.33	0.30	0.33	0.26	
SBQ43	Tries to help someone hurt	0.35	0.30	0.33				0.26	
SBQ45	Spontaneously helps	0.37	0.32	0.34	0.34	0.35	0.37	0.35	0.32
SBQ46	Comforts upset child	0.34	0.31	0.31		0.33	0.33		
SBQ49	Shares things	0.33		0.32		0.31	0.36	0.30	
SBQ50	Encourages bullying	0.38	0.38	0.35	0.36	0.37	0.45	0.46	0.38
SBQ51	Tries to dominate	0.37	0.39	0.38	0.35	0.38	0.44	0.43	0.41
SBQ52	Scares other children	0.41	0.44	0.43	0.35	0.41	0.49	0.47	0.40
SBQ53	Aggressive if teased	0.51	0.52	0.49	0.45	0.48	0.48	0.48	0.45
SBQ55	Aggressive if contradicted	0.48	0.50	0.48	0.43	0.44	0.49	0.47	0.45
SBQ54	Aggressive if something taken	0.47	0.52	0.50	0.41	0.45	0.48	0.48	0.41

Note. Not showing loadings < |0.3|, *SBQ* social behavior questionnaire

extraction of the *p*-factor across a range of item sets, its content (or high loading items) is inexorably linked to the range and specific content of that set. In previous studies, the general factors have been variously tilted towards thought disorder (Caspi et al. 2014; Laceulle et al. 2015); generalised anxiety disorder/major depressive disorder (Lahey et al. 2011; Tackett et al. 2013); and distress (Lahey et al. 2012). In the current study, the *p*-factor was more heavily defined by ADHD and aggressive behaviours whereas pro-social and internalising behaviours tended to have much smaller, often < |0.3|, *p*-factor loadings.

In terms of the importance of the *p*-factor, the minimum ω_h magnitude for the factor solution judged to be the best representation of the data was 0.53 across the eight time points included in the study. This suggests a moderately strong *p*-factor in the SBQ, placing it between the strength of the highly

Table 3 Schmid-Leiman loadings for specific factors over time

Item	Abbreviated content	Measurement wave							
		Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
Aggression									
SBQ10	Impulsive		0.35	0.34					
SBQ11	Impatience with turn-taking		0.35	0.35					
SBQ25	Stealing							0.36	
SBQ26	Destroys own things	0.31			0.31	0.30			0.36
SBQ27	Disobedient	0.41	0.36	0.38	0.41	0.44	0.35	0.37	0.45
SBQ30	Ignores you		0.30	0.32		0.33	0.31		
SBQ31	Destroys others' things	0.41	0.32	0.34	0.41	0.35		0.38	0.43
SBQ32	Lies and cheats	0.30	0.34	0.33	0.30	0.34	0.30	0.36	0.35
SBQ33	Gets into fights	0.57	0.51	0.56	0.57	0.48	0.51	0.57	0.56
SBQ34	Physically attacks	0.63	0.57	0.60	0.63	0.54	0.56	0.60	0.59
SBQ35	Kicks, bites, hits	0.61	0.56	0.59	0.61	0.50	0.52	0.62	0.54
SBQ36	Cruel, bullies	0.53	0.53	0.55	0.53	0.57	0.52	0.60	0.56
SBQ37	Threatens	0.58	0.56	0.60	0.58	0.59	0.58	0.64	0.54
SBQ50	Encourages bullying	0.46	0.44	0.39	0.46	0.53	0.45	0.55	0.60
SBQ51	Tries to dominate	0.50	0.45	0.45	0.50	0.57	0.48	0.54	0.61
SBQ52	Scares other children	0.58	0.50	0.48	0.58	0.61	0.56	0.63	0.43
SBQ53	Aggressive if teased	0.50	0.52	0.45	0.50	0.48	0.41	0.43	0.48
SBQ55	Aggressive if contradicted	0.53	0.53	0.47	0.53	0.51	0.46	0.44	0.45
SBQ54	Aggressive if something taken	0.50	0.51	0.46	0.50	0.46	0.43	0.43	0.54
ADHD									
SBQ10	Impulsive	0.38	0.33	0.32	0.32	0.35	0.34	0.42	0.36
SBQ11	Impatience with turn-taking	0.35	0.30			0.32	0.31	0.46	0.35
SBQ12	Restless/hyperactive	0.52	0.49	0.45	0.43	0.48	0.51	0.54	0.47
SBQ13	Fidgets	0.51	0.48	0.45	0.42	0.46	0.49	0.53	0.46
SBQ14	Can't settle	0.69	0.71	0.65	0.60	0.63	0.71	0.60	0.59
SBQ15	Distractible	0.71	0.74	0.67	0.63	0.65	0.72	0.65	0.64
SBQ16	Can't sustain concentration	0.73	0.75	0.69	0.64	0.66	0.75	0.63	0.62
SBQ17	Inattentive	0.62	0.67	0.60	0.56	0.57	0.64	0.56	0.54
Pro-sociality									
SBQ41	Volunteers to help	0.63	0.66	0.68	0.57	0.68	0.61	0.66	0.67
SBQ42	Tries to stop disputes	0.64	0.64	0.63	0.64	0.70	0.64	0.77	0.74
SBQ43	Tries to help someone hurt	0.75	0.77	0.76	0.71	0.73	0.76	0.78	0.75
SBQ44	Invites bystanders to join game	0.59	0.59	0.49	0.49	0.48	0.42	0.69	0.62
SBQ45	Spontaneously helps	0.70	0.71	0.70	0.64	0.62	0.62	0.69	0.69
SBQ46	Comforts upset child	0.76	0.78	0.78	0.72	0.68	0.74	0.75	0.67
SBQ49	Shares things	0.58	0.58	0.50	0.59	0.49	0.57	0.55	0.55
Internalising									
SBQ02	Nervous, tense	0.30	0.37	0.36	0.35	0.40	0.36	0.42	0.42
SBQ03	Fearful/anxious	0.49	0.55	0.52	0.52	0.56	0.49	0.58	0.58
SBQ04	Worried	0.62	0.70	0.61	0.70	0.68	0.65	0.68	0.70
SBQ05	Unhappy/sad/depressed	0.82	0.83	0.81	0.79	0.79	0.78	0.82	0.83
SBQ06	Not so happy	0.81	0.79	0.82	0.79	0.79	0.81	0.83	0.82
SBQ07	Anhedonic	0.65	0.59	0.65	0.62	0.64	0.63	0.71	0.65
SBQ08	Miserable/ distressed/unhappy	0.71	0.73	0.75	0.72	0.68	0.73	0.79	0.75

Note. Not showing loadings < |0.3|. *SBQ* social behavior questionnaire
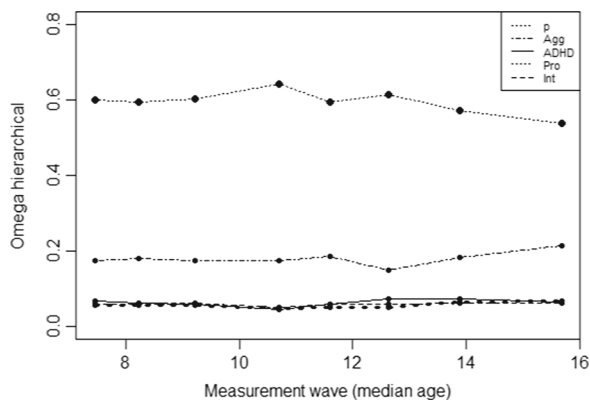
Note. p= p-factor, agg= aggression, ADHD= attention deficit hyperactivity disorder Pro=pro-sociality, Int= internalising.

Fig. 2 Omega hierarchical values across the eight measurement waves for the general and specific factors. *Note. p = p-factor, agg* aggression, *ADHD* attention deficit hyperactivity disorder, *Pro* pro-sociality, *Int* internalising

controversial general factor of personality (GFP) which tends to yield ω_h values around 0.37 and the general factor of cognitive ability (g) which tends to yield ω_h values around 0.74 (Revelle and Wilt 2013). ω_h is, however, dependent on the number and diversity of items analysed, all else being equal increasing with the former and decreasing with the latter. Thus, the strength of the p-factor in the current study should be interpreted in the context of the items from which it was derived which can be argued to cover several distinct domains of common psychopathological symptoms but lacking items from rarer or more severe disorders. For example, the instrument used in the current study does not include any items measuring thought disorder, autism spectrum disorders, many personality disorders, or eating disorders but focuses primarily on internalising and externalising symptoms. Similarly, the range of disorders represented within each specific factor is limited with, for example, phobic, obsessive-compulsive disorder and panic disorder symptoms not represented among the anxiety symptoms. Given that no study to date has included a set of indicators which provides complete coverage of psychopathological behaviours as they are currently defined, an important future direction will be to establish whether the p-factor remains as strong in a more comprehensive item set.

Apparent p-factor strength is also closely linked the methodology used to estimate a p-factor model and the vast majority of studies have used a method which is liable to produce inflated p-factor loadings. In the CFA bi-factor models used in previous studies, it is customary to constrain the majority of cross-loadings to zero; however, in reality a large number of small cross-loadings would be expected both because observed psychopathology symptoms are complex in the sense of reflecting more than one underlying factor and because it is very difficult in practice to design items that are 'pure' measures of only one underlying factor. Constraining cross-

loadings resulting from this kind of complexity to zero forces this covariation to be mediated by other available pathways and is likely to inflate p-factor loadings as a result (Asparouhov and Muthén 2009; Murray and Johnson 2013). An important future direction will, therefore, be to evaluate whether p-factor strengths of the order identified in past studies can be replicated using methodologies such as Exploratory Structural Equation Modelling (ESEM; Asparouhov and Muthén 2009; Morin et al. 2016) and Bayesian Structural Equation Modelling (BSEM; Muthén and Asparouhov 2012) that do not involve the unrealistic assumption of a majority of cross-loadings being zero.

A second question of interest is where in the latent structure of psychopathology, symptoms related to these additional disorders might optimally fit. A few studies have begun to ask questions of this sort. For example, Noordhof et al. (2015) integrated autism spectrum disorder (ASD) symptoms into a bi-factor psychopathology model. They found that the optimal factorial representation of their set of symptoms included a specific factor of ASD that was distinct from the specific externalising, internalising factors and attention/orientation problems factors.

The primary focus of our study was, however, whether p-factor strength—as an indicator of the general covariance among psychopathological behaviours—changed over time. Results suggested that p-factor strength varied within a relatively narrow range and did not systematically increase or decrease over time. This consistency of p-factor strength identified is a potentially important finding because the period covered by the study (i.e., entry to school to adolescence) is a time of significant social, biological and psychological change and development; change that one might expect to be some way reflected in patterns of symptom inter-relations (Cicchetti and Rogosch 2002; Nagin and Tremblay 1999). It is also during this time period that many psychopathological disorders commonly have their onset (e.g., the median ages of onset for anxiety disorders and impulse control disorders are around 11 years of age and others—especially involving delinquency – are quite specific to this time period; Kessler et al. 2005; Ormel et al. 2015). However, the pattern of variation in p-factor strength observed in the current study suggests that a simple version of dynamic mutualism on the one hand and p-factor differentiation on the other does not characterise co-morbidity development in this period. It cannot be ruled out that these kinds of processes characterise p-factor development early on with relative stability in symptom covariance thereafter. High levels of psychopathological co-morbidity are already evident by childhood and early adolescence (e.g., Lahey et al. 2004), therefore, it may be necessary to go further back in development to understand if patterns of co-morbidity are laid down very early or require some time to grow and crystallise. Another possibility that could account for our results is possible diminishing reliability of teacher

reports masking any increase in *p*-factor and/or specific factor strength (e.g., Edelbrock et al. 1985).

The consistency of *p*-factor strength was in the context of low wave-to-wave stability in *p*-factor scores, highlighting the dissociation between stability in levels of general psychopathology and consistency in symptom-general covariance. The stability of *p*-factor scores provides some preliminary insights into the extent to which general psychopathology can be considered a trait-like versus state-like phenomenon. The standardised attenuation-corrected autoregressive coefficients for the *p*-factor ranged from 0.12 to 0.43. It is instructive to consider how the stability of the *p*-factor compares to other developmentally changing constructs. It is clear that its stability falls far short of that of many prototypical psychological traits. For example, intellectual ability is highly stable by middle childhood and can, for example, be expected to correlate between the ages of 10 and 12 at around 0.70 (Bartels et al. 2002; Tucker-Drob and Briley 2014); much larger than the correlation of 0.26 that described the stability of *p*-factor scores around this age and time span. This suggests that, consistent with patterns observed in many specific symptoms of psychopathology, general psychopathology manifests in an episodic fashion. That is, periods of experiencing high (or low) levels of general psychopathology do not have a strong tendency to persist beyond time spans of a year or more.

One possibility is that individuals have a relatively stable pre-disposition towards experiencing a certain set of core symptoms which create secondary issues; however, the composition and relative prominence of these secondary issues may change over time depending on current social circumstances. An individual with a tendency towards depressive mood states may, for example, behave irritably and aggressively during their childhood and adolescence but transition to 'self-medicating' substance abuse as the consequences of aggression and the availability of drugs and alcohol increase. Although their tendency to experience co-occurring problems may change little, the actual manifestation of secondary problems could result in an unstable *p*. In partial support of this hypothesis, the specific factor stabilities for a given measurement interval almost always exceeded that of the corresponding *p*-factor stability. Whatever the precise mechanism underlying this pattern, it would suggest that to the extent that psychopathological symptoms are stable over time, this owes more to the stability of narrower trans-diagnostic factors such as ADHD or Internalising than to a broad, all-encompassing *p*-factor.

However, it should be noted that in estimating stability we relied on two-step approach of first estimating factor scores and then fitting an autoregressive model to these scores. A more optimal estimate of *p*- and specific factor stability may be attained using an explicit measurement model in a CFA (or exploratory structural equation modelling) framework once there is more empirical evidence to guide the optimal factor

structure of the SBQ and other inventories used to measure a *p*-factor.

It would also be of interest to extend observations into adulthood to ascertain if and how general co-morbidity levels and patterns changes across the entire lifespan. General co-morbidity may be affected by both developmental processes such as maturation and aging as well as significant life events and transitions (e.g., leaving school, entering the workforce or getting married). For example, there is evidence that neuroticism decreases with age (Roberts et al. 2006) and given that this trait has linked to difficulties in domains across a range of psychopathological disorders (e.g., Barlow et al. 2014), one might predict that the general psychopathology levels and possibly covariance would show a corresponding decrease. Although *p*-factor studies have been conducted in both childhood and adulthood and reached similar conclusions regarding the presence of substantial general co-morbidity as well as more specific co-morbidity (e.g., Caspi et al. 2014; Patalay et al. 2015), no study has as yet directly compared psychopathology structure across the childhood and adolescence within a *p*-factor framework.

Ultimately, the *p*-factor approach provides only a very general summary of the relation between psychopathological indicators and the analysis of the specific causal pathways linking the constituent disorders remains an important complement to this approach. In this study, the *p*-factor was assessed cross-sectionally at each time point and results, therefore, do not directly inform about symptom continuity, persistence and recurrence or about differential symptom trajectories over time.

It is also necessary to be cautious about reifying the *p*-factor. In the absence of other evidence, it should be considered only a statistical summary of the covariance among psychopathological symptoms, the cause(s) of which is yet to be definitively determined. Answering this question represents a significant challenge if research into general factors in other domains serves as any guide: more than 100 years since Spearman (1904) first described the positive manifold ('*g*') in cognitive ability research, the nature of *g* remains unclear. The models developed in the course of attempting to understand the nature of *g* may be instructive in attempting to unravel the mystery of the *p*-factor. For example, models such as Thompson's bonds model (see Bartholomew et al. 2009) or the dynamic mutualism model discussed in the current study (van Der Maas et al. 2006) provide alternative explanations to the traditionally dominant interpretation of factor models as capturing a underlying latent causal factors.

Relatedly, the bi-factor model is only one technique by which symptom covariance can be modelled and others, for example, network analysis (Borsboom et al. 2011; Borsboom and Cramer 2013; Cramer et al. 2010) provide useful complementary frameworks for developing and testing hypotheses regarding the nature and cause of psychopathological co-

morbidity. For example, while a bi-factor approach may foster thinking about possible common causes for multiple symptoms, a network approach encourages thinking about linkages between symptoms. In the network perspective, particularly influential symptoms or behaviours in a broader network may be identified on the basis of the number and strength of their linkages with others. Similarly, a network perspective may provide a useful framework for the development testing of interventions of designed to break maladaptive linkages between symptoms. Finally, network analysis naturally lends itself to measuring and testing hypotheses about linkages between symptoms within people over time, allowing for the possibility that there may be individual differences in patterns of symptom inter-relations.

Limitations

It is possible that the general co-morbidity captured by the p-factor is at least to partly artifactual. It has been suggested, for example, that implicit theories about psychopathology (i.e., presumptions about which symptoms tend to go together) could lead to inflated symptom inter-correlations. However, a response to this criticism is that implicit theories of psychopathology would tend to hold that specific subsets of, but not *all* psychopathological symptoms tend cluster together (Lahey et al. 2012). Similarly, it has been noted that the p-factor may represent an 'evaluation bias' reflecting individual differences in the tendency to answer questions in a negative or pessimistic manner (e.g., Ye 2009). Some previous studies have attempted to address the possibility that these kinds of rater effects are responsible for the p-factor and demonstrated, for example, that similar results are obtained irrespective of whether self- or informant ratings are used (Tackett et al. 2013) and that substantive criterion associations can be found even when different raters are used for psychopathology symptoms and the criterion measures (Lahey et al. 2015).

Another source of artifact is item context effects (i.e., when responses to items affect responses to subsequent items, artificially inflating their similarity). It has been suggested that the effects of item context will be greatest when items measuring the same construct are presented together or in such a way that it is obvious to the respondent that they are intended to measure the same construct. In the current study items were presented together in a list organised according to the domains outlined in the Measures section. These domains do not correspond exactly to the specific factors extracted in the p-factor analyses but are similar, therefore, correlations both within dimensions and across the entire inventory could have been inflated. However, previous research has suggested that the practical importance of these effects is likely to be minimal (Harrison et al. 1996). Another source of inflated inter-correlations is a

common method effect due to the fact that all but the pro-sociality items were keyed in the same direction and assessed by the same method (i.e., pencil and paper questionnaire; Podsakoff et al. 2012). Future research using multi-trait multi-method based estimates of p-factor strength would help to determine the extent to which this represents an important source of common symptom variance when attempting to measure the p-factor.

Nonetheless, while it is important to acknowledge that the p-factor strength is possibly overstated due to measurement issues, taking into consideration the replicability of the p-factor across different methods of assessment, samples and statistical controls for measurement artifacts, together with the evidence for a range of external variables that could contribute to a shared variance among diverse psychopathological symptoms, it seems unlikely that the general co-morbidity that the p-factor captures is entirely a measurement artifact. Moreover, in regards to the results of the current study, there is no reason to think that these potential artifacts would vary systematically over time and thus mask either p-factor differentiation or growth. A more important limitation in this respect is that to facilitate comparisons across time, we focussed on the set of items that were common across all measurement waves. These items were administered across all waves because they were deemed developmentally appropriate at all studied ages. It is possible that by virtue of this fact, they show higher stability of co-morbidity patterns than would symptoms tend to manifest only earlier or later in development. However, it would not have been possible to include these symptoms in the current study because it would have rendered the omega hierarchical values across waves non-comparable.

Finally, focussing on a questionnaire-based estimate of psychopathology symptoms in a normative sample rather than clinical diagnoses and/or a clinically diagnosed sample has both advantages and disadvantages. In using symptom-level estimates measured in this way we made the assumption that psychopathological symptoms can be conceptualised as continua along which there is meaningful variation at both the clinical and sub-clinical level. If this assumption holds then our measurement approach can capture greater variation in symptoms and avoid the problems associated with artificial dichotomisation or of range restriction due to focussing on clinically diagnosed individuals (e.g., Maxwell and Delaney 1993; Murray et al. 2014a). Furthermore, by focussing on the symptom rather than the diagnosis level, the issue of 'artifactual co-morbidity' (i.e., co-morbidity due to the fact that different clinical diagnoses have some symptoms in common; Rutter 1997) can be avoided. However, the possibility that clinical and sub-clinical levels of some psychopathological disorders have

qualitatively different features has not been definitively ruled out and some have expressed concerns about the meaningfulness and measurement challenges of capturing clinical traits in non-clinical populations (e.g., see Reise and Waller 2009; Murray et al. 2014b).

Conclusions

Much of the variance in psychopathological behaviours is shared with other psychopathological behaviours and can be represented as a p-factor. The extent to which there is general covariance amongst psychopathological behaviours as measured by p-factor strength remains similar from when children enter school through to adolescence. This suggests that the interactions among psychopathological symptoms are not characterised by a simple dynamic mutualism process during this time period, nor are they characterised by a process of p differentiation whereby the manifestation of a general liability towards psychopathology becomes increasingly domain-specific.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 397–438.

Averdijk, M., Zirk-Sadowski, J., Ribeaud, D., & Eisner, M. (2016). Long-term effects of two childhood psychosocial interventions on adolescent delinquency, substance use, and antisocial behavior: a cluster randomized controlled trial. Manuscript submitted for publication.

Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R., & Ellard, K. K. (2014). The nature, diagnosis, and treatment of neuroticism back to the future. *Clinical Psychological Science, 2*, 344–365.

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics, 32*, 237–249.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review, 116*, 567–579.

Behar, L., & Stringfield, S. (1974). *Manual for the preschool behavior questionnaire*. Durham: Lenore Behar.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121.

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS One, 6*, e27407.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2*, 119–137.

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology, 70*, 6–20.

Cole, D. A., Peeke, L. G., Martin, J. M., Truglio, R., & Seroczynski, A. D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology, 66*, 451–460.

Cramer, A. O., Waldorp, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences, 33*, 137–150.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development*, 265-275.

Gignac, G. E. (2014). Dynamic mutualism versus g factor theory: an empirical test. *Intelligence, 42*, 89–97.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: LEA.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*, 430–450.

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes, 68*, 246–261.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41–54.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*, 617–627.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2*, 111–133.

Laceulle, O. M., Vollebergh, W. A., & Ormel, J. (2015). The structure of psychopathology in adolescence: Replication of a general psychopathology factor in the TRAILS study. *Clinical Psychological Science*, epub ahead of print.

Lahey, B. B., Applegate, B., Waldman, I. D., Loft, J. D., Hankin, B. L., & Rick, J. (2004). The structure of child and adolescent psychopathology: generating new hypotheses. *Journal of Abnormal Psychology, 113*, 358–385.

Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D., & Rathouz, P. J. (2011). Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of General Psychiatry, 68*, 181–189.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology, 121*, 971–977.

Lahey, B. B., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry, 56*, 415–422.

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: a cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology, 40*, 677–692.

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

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*, 181–190.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: LEA.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*, 116–139.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*, 407–422.

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014a). Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the Autism Spectrum Quotient Short Form. *Journal of Autism and Developmental Disorders, 44*, 55–64.

Murray, A. L., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014b). Are we under- estimating the association between autism symptoms?: The importance of considering simultaneous selection when using samples of individuals who meet diagnostic criteria for an Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 44*, 2921–2930.

Murray, A. L., Eisner, M. P., & Ribeaud, D. (2016). The need for dimensional, trans-diagnostic measures of childhood and adolescent psychopathology: an analysis of the Social Behavior Questionnaire. Manuscript submitted for publication.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods, 17*, 313–335.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181–1196.

Noordhof, A., Krueger, R. F., Ormel, J., Oldehinkel, A. J., & Hartman, C. A. (2015). Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. *The TRAILS study. Journal of Abnormal Child Psychology, 43*, 577–587.

Ormel, J., Raven, D., van Oort, F., Hartman, C. A., Reijneveld, S. A., Veenstra, R., & Oldehinkel, A. J. (2015). Mental health in Dutch adolescents: a TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. *Psychological Medicine, 45*, 345–360.

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry, 207*, 15–122.

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. *Annual Review of Psychology, 63*, 539–569.

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for statistical computing. **http://www.R-project.org**.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27–48.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modelling: a bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26.

Revelle, W., & Wilt, J. (2012). On when a factor is a general factor: Presented at the European Association for Personality Psychology experts meeting on the structure of personality, Dubrovnik, Croatia.

Revelle, W., & Wilt, J. (2013). The general factor of personality: a general critique. *Journal of Research in Personality, 47*, 493–504.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 1–25.

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: preliminary findings. *Journal of Child Psychology and Psychiatry, 8*, 1–11.

Rutter, M. (1997). Comorbidity: concepts, claims and choices. *Criminal Behaviour and Mental Health, 7*, 265–285.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*, 201–292.

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. *Psychological Medicine, 45*, 1483–1493.

Tackett, J. L., Lahey, B. B., Van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology, 122*, 1142–1153.

ten Berge, J. M. F., Krijnen, W. P., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications, 289*, 311–318.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology, 19*, 285–300.

Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: a meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin, 140*, 949–979.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1–67.

van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review, 113*, 842–861.

van Lier, P. A., Vitaro, F., Barker, E. D., Brendgen, M., Tremblay, R. E., & Boivin, M. (2012). Peer victimization, poor academic achievement, and the link between childhood externalizing and internalizing problems. *Child Development, 83*, 1775–1788.

Weir, K., & Duveen, G. (1981). Further development and validation of the prosocial behaviour questionnaire for use by teachers. *Journal of Child Psychology and Psychiatry, 22*, 357–374.

Wright, A. G., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: toward an

expanded quantitative empirical model. *Journal of Abnormal Psychology, 122*, 281–294.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): the role of wording effects. *Personality and Individual Differences, 46*, 197–201.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher- order factor model and the hierarchical factor model. *Psychometrika, 64*, 113–128.