Development and Psychopathology **30** (2018), 1371–1387

doi:10.1017/S0954579417001602

Do executive functions explain the covariance between internalizing and externalizing behaviors?

ALEXANDER S. HATOUM, SOO HYUN RHEE, ROBIN P. CORLEY, JOHN K. HEWITT, AND NAOMI P. FRIEDMAN
University of Colorado Boulder

Abstract

This study examined whether executive functions (EFs) might be common features of internalizing and externalizing behavior problems across development. We examined relations between three EF latent variables (a common EF factor and factors specific to updating working memory and shifting sets), constructed from nine laboratory tasks administered at age 17, to latent growth intercept (capturing stability) and slope (capturing change) factors of teacher- and parent-reported internalizing and externalizing behaviors in 885 individual twins aged 7 to 16 years. We then estimated the proportion of intercept–intercept and slope–slope correlations predicted by EF as well as the association between EFs and a common psychopathology factor (P factor) estimated from all 9 years of internalizing and externalizing measures. Common EF was negatively associated with the intercepts of teacher-rated internalizing and externalizing behavior in males, and explained 32% of their covariance; in the P factor model, common EF was associated with the P factor in males. Shifting-specific was positively associated with the externalizing slope across sex. EFs did not explain covariation between parent-rated behaviors. These results suggest that EFs are associated with stable problem behavior variation, explain small proportions of covariance, and are a risk factor that that may depend on gender.

Factor-analytic methods have specified two factors that account for shared variation across different problem behavior symptoms (e.g., Lilienfeld, 2003). Internalizing behavior, including anxiety and depression, captures the tendency to withdraw, or internalize, distress. Externalizing behavior, including delinquency and antisocial behaviors, captures the tendency to express outward, or externalize, distress. However, these internalizing and externalizing factors significantly covary (e.g., Lilienfeld, 2003). Although researchers have examined whether personality and behavioral liabilities explain this covariance (e.g., Rhee, Lahey, & Waldman, 2015), few have investigated the role of cognitive abilities, specifically executive functions (EFs). In this study, we examined the relations between multiple separable EFs and trajectories of internalizing, externalizing, and their covariance across childhood and adolescence, including whether EFs' associations with internalizing and externalizing behavior is due to a common liability, or general psychopathology factor. In addition, we investigated whether these relations differed for boys and girls.

Covariance Between Internalizing and Externalizing Behavior

Latent factors can be used to capture common variance across individual behaviors (such as symptoms of disorders; e.g., Kroes et al., 2002), or across multiple disorder diagnoses (e.g., Cosgrove et al., 2011). Although internalizing and externalizing factors are separable, recent research has incorporated a common factor, or P factor, that can be extracted across measures of child, adolescent, and adult problems (Caspi et al., 2014; Laceulle, Vollebergh, & Ormel, 2015; Martel et al., 2017; Snyder, Young, & Hankin, 2017). Individuals with higher co-occurring internalizing and externalizing disorders or symptoms often have more debilitating symptoms and impairments (Cerdá, Sagdeo, & Galea, 2008), and this pattern extends to covarying clinical diagnoses (Cosgrove et al., 2011). Thus, it has become an especially important challenge for researchers to account for this covariance across internalizing and externalizing behavior.

Covariance between internalizing and externalizing behaviors extends to patterns across development as well. A popular approach to studying such development uses latent variable growth models, in which individual differences in trajectories can be described with latent intercept and slope factors. In typical parameterizations, the intercept factor captures individual differences in the first time point, and the variation in other time points shared with that time point. The slope captures individual differences in the change across the time points included. Keiley, Bates, Dodge, and Pettit (2000) argued that modeling the trajectories offers more information on the nature of the behavior, including how it grows, diminishes, and for purposes of our study, how different behaviors relate to these patterns across time. Across constructs, both the stability factors (intercepts) and the change factors (slopes) of one behavior tend to correlate significantly

This research was supported by NIH Grants MH063207, AG046938, HD010333, and MH016880.

Address correspondence and reprint requests to: Alexander S. Hatoum, Institute for Behavioral Genetics, 447 UCB, University of Colorado, Boulder, CO 80309; E-mail: alexander.hatoum@colorado.edu.

with those of the other (Keiley et al., 2000), suggesting that the covariance between internalizing and externalizing behavior extends past one time point. That is, there is covariance not only between the stable variances for these behaviors but also between their changes across time.

Past studies have included covariates to distinguish between disorders and explain common liabilities. There are three general categories under which covariates fall (Weiss, Susser, & Catron, 1998): common features that distinguish both internalizing and externalizing from normality (which may also relate to the P factor); broadband features that distinguish internalizing and externalizing from one another; and specific features within particular areas of internalizing and externalizing (e.g., conduct problems vs. substance use). For example, neuroticism seems to act as a common feature (Hink et al., 2013), whereas disinhibition seems to be a broadband feature that is unique to externalizing behaviors (Krueger & Tackett, 2003).

Executive Functioning

EFs are candidates for common features, because they have been proposed as transdiagnostic features of psychopathology (Goschke, 2014; McGrath et al., 2016; Nolen-Hoeksema & Watkins, 2011). EFs are high-level cognitive abilities that regulate goal-directed behaviors. The term EF has been used to describe a number of abilities, including inhibiting responses, ignoring distraction, switching between tasks, working memory maintenance, and updating, planning, and verbal fluency (Diamond, 2013). Although these abilities are separable (Miyake et al., 2000), they do share variance, and a common EF factor has been used to account for correlations across tasks or latent variables (e.g., Miyake & Friedman, 2012). This common EF factor is separable from general cognitive ability/intelligence (see Friedman & Miyake, 2017, for an in-depth discussion).

Studies focused on particular internalizing and externalizing disorders suggest that EFs are related to both of these constructs. In their review of the literature, Sergeant, Geurts, and Oosterlaan (2002) found that EFs (measured with individual tasks tapping inhibition, working memory, set shifting, planning, and fluency) are negatively related to disorders that load on the externalizing factor, including attention-deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder. Meta-analyses also suggest that EF tasks (spanning volition, planning, purposeful action, and effective performance) are significantly negatively related to antisocial personality disorder (Morgan & Lilienfeld, 2000; Oglivie, Steward, Chan, & Shum, 2011). With respect to internalizing disorders, a recent meta-analysis (Snyder, 2013) found that individuals with major depression had deficits on multiple kinds of EF tasks (tapping inhibition, working memory set shifting, and fluency). Another meta-analysis examining the same EFs found broad EF deficits in obsessive–compulsive disorder (Snyder, Kaiser, Warren, & Heller, 2015). A recent review (Snyder, Miyake, & Hankin, 2015) concluded that deficits in multiple EFs were broadly associated with psychopathology. Taken together, these reviews and meta-analyses are consistent with the conclusion that multiple forms of psychopathology are negatively associated with deficits in a general EF factor.

In addition, various EF tasks have been related individually to internalizing and externalizing behaviors across development. Riggs, Blair, and Greenberg (2004) found that first- and second-grade children's sequencing (as measured by the trail-making test) and inhibition (as measured by the Stroop task) negatively predicted later teacher-rated internalizing and externalizing behaviors. These authors focused on these EF abilities because both are aspects of cognitive forethought; they hypothesized that EF exerts influence over internalizing and externalizing behaviors by discouraging the activation of behaviors with negative future consequences. Furthermore, Hughes and Ensor (2011) found that from ages 4 to 6, the growth (latent slope factors) of inhibition and working memory were negatively related to multiple behavior problems, including internalizing and externalizing. Using a subset of the same data set used in the current study, Young et al. (2009) found that externalizing behavior, measured with a behavior disinhibition latent variable (with factor loadings for substance use, conduct disorder, attention-deficit/hyperactivity disorder symptoms, and novelty seeking personality) at ages 12 and 17 years significantly negatively correlated with a response inhibition latent variable at age 17. Taken together, this research establishes that a number of EF tasks tapping correlated but separable EFs are negatively associated with psychiatric behaviors.

Finally, some studies have linked general EF to a P factor. Martel et al. (2017) reported that a global EF factor (including a conflict control task, a go/no-go task, digit span, Corsi blocks, and a time anticipation task) significantly predicted a P factor (based on parents' reports on diagnostic questions for their children, ages 6 to 12 years). Caspi et al. (2014) also found in adults assessed longitudinally from ages 18 to 38 years that individual EF tests (Trail Making B, Wechsler Memory Scale—III Mental Control, and Cambridge Neuropsychological Test Automated Battery Rapid Visual Information Processing: A-Prime) were negatively correlated with a P factor.

The Current Study

As this brief review indicates, a number of studies support the hypothesis that EFs, broadly considered, are associated with individual differences in internalizing and externalizing behaviors, as well as their covariance. However, as Snyder, Miyake, et al. (2015) discussed, the bulk of the clinical literature linking psychopathology to EFs has not fully connected to recent theoretical and methodological advances in cognitive psychology. Specifically, many clinical studies focus on individual neuropsychological tasks, which is problematic for two main reasons (Miyake et al., 2000). First, individual tasks are particularly impure measures of the targeted EF;

because EF tasks necessitate acting on other cognitive processes, variation in task performance can be due to these other cognitive processes in addition to the EF of interest. Thus, if psychopathology is associated with an EF task, that association could reflect non-EF cognitive processes rather than the EF of interest. Second, as discussed earlier, there are multiple correlated but separable EFs (Friedman & Miyake, 2017). Thus, a wealth of research broadly agrees on a model in which EFs share some common cognitive processes, but also in some cases include EF-specific processes (i.e., specific to set shifting or working memory updating; see Friedman & Miyake, 2017, for a review). An individual EF task will tap both common and specific EF processes, in addition to non-EF cognitive processes, making associations difficult to interpret. Thus, an investigation of the source of common variance in behavior problems might benefit from incorporating a well-validated model of the multicomponent structure of EFs (Snyder, Miyake, et al., 2015).

In this study, we examine how multiple EFs relate to trajectories and commonality of internalizing and externalizing behaviors with the unity/diversity framework (Friedman & Miyake, 2017; Friedman et al., 2008; Miyake & Friedman, 2012), which addresses both the task impurity and the multidimensionality problems associated with individual EF tasks. Specifically, this framework describes the relations among three of the most commonly studied EFs at the level of latent variables: response inhibition, working memory updating, and set shifting (Friedman & Miyake, 2017). Latent variables extract common variance across multiple measures; when measures are selected such that they share the EF of interest but differ in non-EF requirements, then the latent variables provide purer measures of the underlying constructs that are free from random measurement error (Bollen, 1989). Thus, the tasks included in the unity/diversity framework were selected to tap one the three most commonly studied EFs (response inhibition, working memory updating, and mental set shifting), but to differ in their lower level cognitive requirements. For example, the response inhibition tasks required inhibiting dominant eye movements, word reading, or semantic categorization responses. Moreover, these tasks were selected to be reliable measures of individual differences that did not tap multiple separable EFs (i.e., were not measures of potentially more complex EFs like planning; Miyake et al., 2000).

At the latent-variable level, these three abilities are correlated but separable (Friedman et al., 2008; Miyake et al., 2000), with latent variable correlations ranging from .38 to .79 in this sample. The unity/diversity framework (Figure 1) captures this structure with a common EF latent variable that predicts all nine EF tasks, and orthogonal updating-specific and shifting-specific factors that capture remaining correlations among the three updating and three shifting tasks, respectively, once the common EF variance is removed. There is no "inhibiting-specific" factor because the common EF factor explains all the correlations among the inhibiting tasks; in other words, the common EF factor is isomorphic with the response inhibition factor, a consistent finding across several independent studies (Friedman & Miyake, 2017; Miyake & Friedman, 2012).

This parameterization is known as a bifactor model. (The term bifactor does not refer to the number of factors; it refers to the structure of the model, which is orthogonal common and specific components with complex loadings, rather than a hierarchical structure.) The bifactor parameterization has several advantages over the correlated factors model. Most important, it captures what is common across multiple EFs with a latent variable that can be related to external correlates, rather than having this common variance represented by the correlations among the factors. In prior work, we have found that this common variance (vs. the specific factors) is the most related to a range of behavior problems (for reviews, see Miyake & Friedman, 2012; and Herd et al., 2014). Moreover, Snyder, Miyake, et al.'s (2015) review also suggested that it is this common EF factor that is transdiagnostic. The

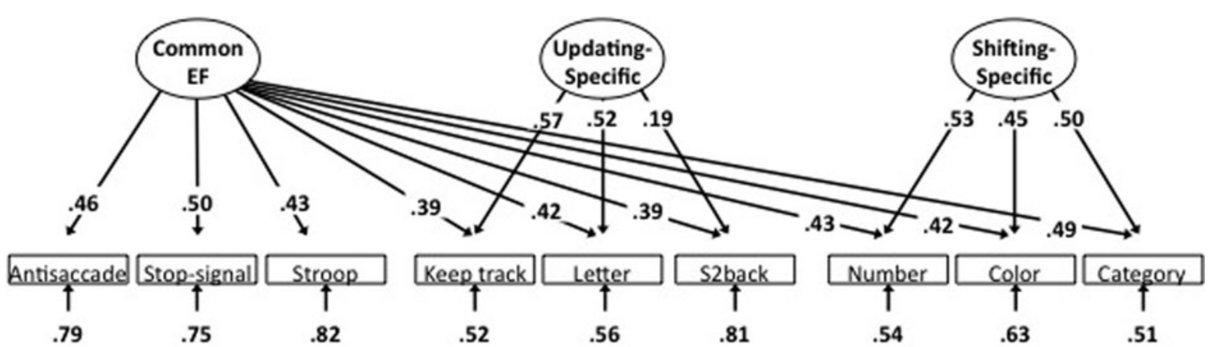

Figure 1. Unity/diversity model of executive functions (EFs), with parameter estimates for the current sample. A common EF factor accounts for shared variation across all nine tasks (including response inhibition, working memory updating, and set shifting tasks), and orthogonal updating-specific and shifting-specific factors account for remaining covariances among the updating and shifting tasks, respectively. There is no inhibiting-specific factor because the common EF factor accounts for all the covariances among the inhibiting tasks. Letter, letter memory; S2back, spatial 2-back; number, number–letter; color, color–shape; category, category switch. From "Unity and Diversity of Executive Functions: Individual Differences as a Window on Cognitive Structure," by N. P. Friedman and A. Miyake, 2017, *Cortex, 86*, 186–204. Copyright 2017 by Elsevier Science. Adapted with permission.

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

common EF factor is thought to capture the ability to actively maintain and manage goals, and use those goals to bias ongoing processing (Friedman & Miyake, 2017; Miyake & Friedman, 2012). This goal management is a general requirement of all EF tasks, and may be particularly important for response inhibition tasks, in which weak goal representations may allow more dominant responses to take over.

In addition, because the specific factors are orthogonal to the common factor and each other, we can examine whether they are independently associated with internalizing and externalizing behaviors without problems due to multicollinearity. The shifting-specific factor is thought to reflect the speed with which goals can be replaced (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Past research has suggested positive associations between the shifting-specific factor and behavioral disinhibition, attention problems, and lower self-restraint (Herd et al., 2014). This pattern of negative association with common EF but positive association with shifting-specific may reflect a stability–flexibility trade-off (Friedman & Miyake, 2017) whereby weak goal maintenance impairs overall performance (lower common EF) but makes it easier to shift to a different goal (better shifting specific). Past research suggests few relations with the updating-specific factor, which is thought to tap aspects of working memory gating (i.e., by the basal ganglia) and potentially memory retrieval (Friedman & Miyake, 2017; Miyake & Friedman, 2012).

The nine-task EF battery was assessed at age 17 years. Similar to past research on internalizing and externalizing behavior, we first related EF factors to trajectories of behavior problems, using parent and teacher ratings from ages 7 to 16 years. In line with recent research on behavior problems, we then related the EF factors to a P factor that utilizes all 9 years of both internalizing and externalizing symptoms.[1] Our primary interest in these models is whether one or more EF components predicts both internalizing and externalizing behaviors, explaining some proportion of their covariation, and whether this EF factor then relates to the P factor. Because the literature reviewed earlier indicates that multiple EFs are related to internalizing and externalizing behaviors, we hypothesize that it will be the common EF factor that underlies at least some of their covariance, in both the bivariate growth models and the P factor model. We may also find a relation with the shifting-specific factor, but in the opposite direction, based on prior research with this model suggesting that behavior problems are sometimes associated with better shifting-specific abilities (Herd et al., 2014).

1. We analyze parent and teacher ratings separately for several reasons. First, teacher and parent reports often show low correlations for internalizing and externalizing data (Achenbach, McConaughy, & Howell 1987), which is consistent with our data (correlation matrices available upon request). Second, the lack of overlap is thought to reflect the raters' unique perspectives and context-dependent experiences with the children (Arseneault et al. 2003; Derks, Hudziak, Beijsterveldt, Dolan, & Boomsma, 2004). Third, multitrait–multimethod techniques have been shown to converge on poor solutions for internalizing and externalizing (Cole, 1987), and could reflect other sources of variance (Marsh, 1989).

Finally, we explore sex differences in EFs' relations to these behaviors. Males tend to show higher levels of externalizing behavior, whereas females tend to show higher levels of internalizing behavior (Keiley, Lofthouse, Bates, Dodge, & Pettit, 2003). Because of differences in treatment, expected reactions to the environment, or exposure to sex-specific environments, sex differences may come about due to an interaction between risk mechanisms and particular social cues (Rutter, Caspi, & Moffit, 2003). Therefore, it is possible that risk factors either predispose sexes differently or differentially affect the manifestation of these problems. If so, we might find that EFs show different patterns of relations to these behaviors across sex.

Method

Participants

Participants were 925 individual same-sex twins (468 females, 444 males) from the Colorado Longitudinal twin study (for more information on this sample, see Rhea, Gross, Haberstick, & Corley, 2013) who had data for any measure for at least one time point. Of these participants, 786 had EF data. The combination of behavior problems and EF data resulted in a total sample size of 885 (450 females, 435 males) for the teacher-rated behavior and 912 (468 females, 444 males) for parent-rated behavior.

All protocols for data collection were reviewed by the University of Colorado Institutional Review Board. Informed consent or assent and parental permission were collected from each participant. Participants were given monetary compensation.

Internalizing and externalizing behavior

Parents completed the Child Behavior Checklist (CBCL; Achenbach, 1991a) when their children were ages 7 and 9 to 16 years. The same parents completed ratings for both twins. Teachers completed the Teacher Report Form (TRF; Achenbach, 1991b) when their students were ages 7 to 15 years. The same teachers rated both twins only if twins were in the same classrooms; otherwise, different teachers rated each twin.

The CBCL and TRF were mailed with a packet of questionnaires in the spring of each year (starting in the first grade and going through sophomore year of high school). Parents either sent the TRF to the teacher or gave it to him or her personally. Both the CBCL and the TRF are checklists of problem behaviors, where each item is rated on a scale of 0 = *not true (as far as you know)*, 1 = *somewhat or sometimes true*, and 2 = *very true or often true*. The TRF and CBCL internalizing and externalizing scales are composites of subscales: the externalizing scale is composed of the aggressive and delinquent subscales totaling 34 items for teachers and 33 items for parents (highest possible scores of 68 and 66, respectively); the internalizing scale is composed of the anxious–

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

depressed attachment style, somatic complaints, and social withdrawal scales, with a total of 35 items for teachers and 31 for parents (highest possible score of 70 and 62, respectively). These scales show concurrent validity when related to other measures of externalizing and internalizing psychopathology (Cohen, Gotlieb, Kershner, & Wehrspann, 1985).

EF

The EF battery was administered at approximately age 17 years ($M = 17.3$, $SD = 0.6$, range = 16.5–20.1). The nine computerized EF tasks were fully described by Friedman et al. (2008). Here we summarize their basic requirements.

The inhibiting tasks required stopping automatic or dominant responses. In the antisaccade task, participants had to resist the reflexive tendency to saccade toward a cue that briefly flashed on one side of the screen, instead immediately saccading to the opposite side of the screen in time to see an arrow that briefly appeared before being masked. The dependent measure was the proportion of correct identifications of the arrow's direction (out of 90). In the stop-signal task, individuals categorized words as animals or not as quickly as possible, except when they heard a signal on 25% of the trials that indicated to withhold their responses. The dependent measure was the stop-signal reaction time, which is the estimated time for the stopping process to finish. In the Stroop task, participants named aloud the colors in which color words, noncolor words, and asterisks were printed, with their response times recorded by a voice key microphone. The dependent measure was average response time to name the colors of incongruent color words minus the average response time to name the colors of asterisks.

The updating tasks required continuously updating working memory with new relevant information, deleting no-longer relevant information when appropriate. In each trial of the keep track task, participants read a series of 15 words belonging to six categories, including multiple exemplars per category, and reported at the end of the list only the last exemplars in two to four prespecified categories (e.g., colors and countries). The dependent measure was the proportion of words correctly recalled (out of 36). In each trial of the letter memory task, participants saw a series of 5, 7, or 9 letters. With each new letter that appeared, they had to say aloud the prior 3 letters; and at the end of the trial, they had to recall the last 3 letters. The dependent measure was the proportion of correctly recalled letters (out of 30) at the end of the trials. In the spatial-2-back task, participants saw 10 squares on the screen darken, 1 at a time. For each trial, they had to respond with a button-press whether the indicated location was the same as the one two trials before. The dependent measure was the proportion of correct responses (yes and no) across four blocks of 25 trials each, counting omissions as errors.

The shifting tasks required rapidly switching between two subtasks. All three tasks used the same two buttons to categorize stimuli based on cues that appeared 150 ms before the stimuli. The cues appeared in a fixed pseudorandom order, such that half the trials were repeat trials (same subtask as the prior trial), and half were switch trials (different subtask as the prior trial). In the number–letter task, participants classified the letter (as consonant or vowel) or number (as odd or even) in a letter–number or number–letter pair (e.g., 7G), depending on whether it appeared in a square in the top or bottom of the screen (the cue was a darkening of the outline of the square). In the color–shape task, the stimulus was a colored rectangle with a shape in it, with the cue above it. Participants classified the shape as a circle or a triangle when the cue was S, and the color as red or green when the cue was C. In the category-switch task, participants categorized words as living or nonliving when the cue above them was a heart, and smaller or bigger than a soccer ball when the cue was a set of crossed arrows. For all three tasks, the dependent measure was the local switch cost, the difference between the average reaction time for switch trials minus the average reaction time for repeat trials (across two blocks with 48 trials each).

Data analysis

All analyses were conducted with Mplus, Versions 7.1–7.3 (Muthen & Muthen, 1998–2014), using the clustering (*type*= *complex*) option. This option uses a weighted likelihood function and a sandwich estimator to obtain a scaled chi-square (χ^2) and standard errors corrected for the nonindependence of individual-level data (in this case, correcting for within-family nonindependence). We assessed model fit with the χ^2 statistic, supplemented with the root mean square error of approximation (RMSEA) and the comparative fit index (CFI). We used an RMSEA value of <0.06 and a CFI value of >0.95 as indications of good fit (Hu & Bentler, 1998). Significance of parameters was determined by z tests based on the ratio of the parameters to their standard errors.

Data transformations. The internalizing and externalizing scores at each year were not normally distributed. Prior work suggests that binning and analyzing such skewed symptom count data as ordinal variables assuming an underlying normal liability distribution results in less biased parameter estimates than transformations (Derks, Dolan, & Boomsma, 2004). Therefore, we binned the scores at each year into four bins: 0, 1–3, 4–10, and >10. These bins were selected prior to analysis to ensure that adequate numbers of subjects would fall into each bin across time points, and the same bins were used for all time points and for both internalizing and externalizing measures (see online-only supplementary Table S.1 for descriptive statistics for each year; *n*s per bin are available upon request). We then analyzed these binned variables as ordinal variables with Mplus using the weighted least squares means and variances adjusted estimator (with delta parameterization). The weighted least squares means and variances adjusted estimator uses pairwise deletion for missing values.

The EF data were identical to those used in past work with this model (Friedman et al., 2016; see online-only supplementary Table S.2 for descriptive statistics). As described by Friedman et al. (2008), these data were normally distributed after within-subject and between-subject trimming, and arcsine transformation of accuracy data. Reaction time data were reversed so that higher scores represent better performance for all measures in the models.

Growth model parameterization. To examine stability and growth of internalizing and externalizing behaviors over time, we estimated latent growth curves.[2] Each growth model (one for internalizing behavior and one for externalizing behavior) included an intercept latent variable and a slope latent variable. The intercept factor had a loading for each time point of 1.0. In the freed-curve model (Bollen & Curran, 2006) we used, the slope factor had a loading for the first time point fixed to zero, the loading for the last time point fixed to 1.0, and the remaining loadings freely estimated. This parameterization fits the growth curve rather than constraining it to follow a particular function. With this parameterization, the intercept captures variation in initial scores and variation at later times that is stable with those initial levels (because all time points load equally on the intercept). The slope captures change from the initial to the final time point; each estimated slope loading represents the proportion of total change.

We compared this freed-curve model to a linear curve for both sexes separately and also for models assuming invariance. For teacher-rated data with separate parameters for males and females, the linear curves did not fit significantly worse than the freed curves, all χ^2 difference (7) < 11.21, $p > .129$. For the teacher-rating model with sex invariance, the freed internalizing curve was not significantly different from a linear curve, χ^2 difference (7) $= 11.22$, $p = .129$, and the freed externalizing curve was marginally different from a linear curve, χ^2 difference (7) $= 12.28$, $p = .092$. In contrast, all but one growth model for the parent-rated data showed significant departures from linearity: internalizing male, χ^2 difference (7) $= 5.78$, $p = .566$; externalizing male, χ^2 difference (7) $= 15.14$, $p = .034$; externalizing female, χ^2 difference (7) $= 26.55$, $p < .001$; internalizing female, χ^2 difference (7) $= 31.96$, $p < .001$; invariant externalizing, χ^2 difference (7) $= 35.98$, $p < .001$; invariant internalizing, χ^2 difference (7) $= 26.00$, $p < .001$. We opted to estimate nonlinear growth curves for all of our final models (teacher and parent ratings) to maintain consistency across teacher and parent ratings. Allowing nonlinearity is also appropriate given prior evidence of nonlinearity in development of problem behaviors (Hinshaw, 2002; Kazdin & Kagan, 1994; Kim & Cicchetti, 2006).

We used the Mplus default for ordinal data and estimated a single set of thresholds for all time points (equated across sex), set the mean of the intercept factor to zero for the first group (in this analysis, females), and freed the mean of the intercept factor in the second group. In the model with both growth curves (internalizing and externalizing), we included time-specific residual correlations (i.e., the age 7 internalizing residual was allowed to correlate with the age 7 externalizing residual, etc., and these residual correlations were allowed to differ across sex).

Sex invariance. We tested for invariance across sex separately for the internalizing and externalizing growth models. We tested whether the loadings could be constrained without detriment in fit to establish metric invariance. Then we tested whether scale factors (analogous to residual variances for continuous measures) could be constrained to equality to establish strict invariance. For teacher ratings, both the externalizing and internalizing models showed strict invariance, both metric χ^2 difference (7) < 7.37, $p > .391$; both strict χ^2 difference (8) < 7.75, $p > .458$. For the parent ratings, only the externalizing model met criteria for strict invariance, metric χ^2 difference (7) $= 2.95$, $p = .890$; strict χ^2 difference (8) $= 2.94$, $p = .938$. The internalizing model failed to meet metric invariance, χ^2 difference (7) $= 14.82$, $p = .038$. Furthermore, we were able to constrain the growth factor variances to be equal across sex in the teacher-rating internalizing and externalizing models, internalizing, χ^2 difference (2) $=$ 1.66, $p = .436$; externalizing, χ^2 difference (2) $= 1.45$, $p =$ .484, as well as the parent-rating externalizing model, χ^2 difference (2) $= 1.52$, $p = .468$. We allowed the covariances to differ by sex to examine whether the sexes showed different amounts of common variance.

Given these results, the final model based on teacher ratings included sex-invariant loadings, scales, and factor variances for both internalizing and externalizing behavior. The model based on parent ratings included sex-invariant loadings, scales, and factor variances for externalizing behavior, but noninvariant parameters for internalizing behavior. In all models, time-specific residual correlations between internalizing and externalizing scores were allowed to vary by sex.

Relation between growth factors and EFs. After estimating a bivariate growth model, we regressed the internalizing and externalizing growth factors on the three EF factors in the bifactor EF model (Figure 1). In the EF model, we constrained the factor loadings and factor variances to be equal across sex, loadings: χ^2 difference (12) $= 9.00$, $p = .487$; variances: χ^2 difference (3) $= 0.98$, $p = .914$, but allowed the intercepts and residual variances for each task to vary between sexes, as invariance tests revealed that these exceptions to measurement invariance were necessary, intercepts: χ^2 difference (9) $= 92.49$, $p < .001$; residuals: χ^2 difference (9) $= 31.08$, $p <$ .001. We also constrained the loadings for the keeptrack and letter memory tasks on the updating-specific factor to be equal to each other to ensure this factor was identified in the multiple-group analysis (given that the loading for spatial

2. There was some attrition in our sample (see online-only supplementary Table S.1). We tested for nonrandom missingness by regressing missingness at the final time point on scores for the first three time points for each problem behavior measure. In all cases, internalizing and externalizing scores at ages 7–9 years did not predict missingness at the final time point (age 15 for parent ratings, 16 for teacher ratings, all *ps* $> .055$).

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

two-back was small, this factor could have been empirically underidentified).

To examine the proportion of the internalizing–externalizing intercept and slope correlations explained by their regressions on EFs, we multiplied the standardized betas from each EF component (i.e., using standard path-tracing rules). We computed these products within the same scripts used to run the models in Mplus, which provided the standard errors and p values for the products that we report.

Finally, to relate EFs to the P factor, we first estimated a general psychopathology factor from all years of data by regressing each time point on the P factor, and estimating residual internalizing and externalizing intercept and slope factors. Because the P factor should capture the common variance across internalizing and externalizing scores, we constrained associations across internalizing and externalizing growth factors to be zero (as shown later), but we did allow the negative intercept with slope correlations. We then correlated the orthogonal components of EFs with the P factor as well as the residual growth factors.

Results

Growth models of internalizing and externalizing behavior

To examine the covariance between the internalizing and externalizing growth factors, we estimated bivariate growth models separately for the parent and teacher ratings. In all models, the females' means for the intercept factors were set to zero and act as a reference for comparison to males. Full correlation matrices are available upon request.

Teacher ratings. The model examining teacher ratings fit well, χ^2 (338) = 354.54, p = .257, CFI = 0.992, RMSEA = 0.011. As shown in Figure 2a, the internalizing intercept mean for males was significant (μ = 0.18, p = .008), indicating that males showed higher initial levels of internalizing behavior than females. The externalizing intercept mean for males was also significant (μ = 0.46, p < .001), consistent with past findings that males have higher levels of externalizing behavior (Keiley et al., 2003). Trajectories appeared nonlinear (see slope loadings in online-only supplementary Table S.3). The means for the slope factors suggested an overall decline in the level of problem behavior from ages 7 to 15. Specifically, for females, the slope mean for externalizing behavior (μ = –0.24, p = .003) was significant, though the slope mean for internalizing behavior did not reach significance (μ = –0.15, p =.060). Males' mean slopes for both internalizing (μ = –0.36, p < .001) and externalizing (μ = –0.31, p = 001) behavior were significant.

All of the growth factors had significant variance (all ps < .027), indicating that there were significant individual differences in these behaviors' initial levels and growth across time. Within each behavior type, intercepts negatively correlated with slopes for females (internalizing r = –.49, p < .001; externalizing r = –.48, p = .001), suggesting that those with higher initial levels showed larger decreases in behavior across time. In males the internalizing intercept–slope correlation did not reach significance (r = –.29, p = .074) but the externalizing intercept–slope correlation did (r = –.33, p = .041), although these correlations were comparable numerically.

Of primary interest to the question of covariance in these behaviors are the cross-trait associations (i.e., the correlations between the internalizing and externalizing intercept factors and the correlations between the internalizing and externalizing slope factors). In line with prior literature suggesting covariance between internalizing and externalizing behaviors, the intercept factors significantly positively correlated in both sexes (females, r = .30, p = .003; males, r = .38, p < .001). However, the slope correlations were not significant in either sex (females, r = .41, p = .087; males, r = .45, p = .088). Neither the correlations of the intercepts, χ^2 difference (1) = 0.256, p = .613, nor the correlations of the slopes, χ^2 difference (1) = 0.011, p = .917, were significantly different across the sexes. In a model in which the intercept and slope correlations were constrained to be equal across sex, both the correlations of the intercepts (r = .34, p < .001) and of the slopes (r = .42, p = .010) were significant.

Parent ratings. The model for parent ratings fit well, χ^2 (320) = 433.47, p < .001, CFI = 0.991, RMSEA = 0.029.[3] Generally, the trajectories followed a similar pattern as those in the model examining the teacher ratings (see online-only supplementary Table S.3). As shown in Figure 2b, males' intercept mean for internalizing behavior was not significant (μ = –0.10, p = .339), whereas males' intercept mean for externalizing behavior was significant and positive (μ = 0.32, p = .001), indicating a higher level of externalizing behavior compared to girls. Both sexes showed a similar decline in internalizing, though the slope mean only reached significance in females (females, μ = –0.18, p = .025; males, μ = –0.17, p = .068). Both sexes showed somewhat larger declines in externalizing behavior (females, μ = –0.50, p < .001; males, μ = –0.45, p < .001).

The four growth factors had significant variances in both sexes (all ps < .021), indicating that there were significant individual differences in both the stability and change in behavior problems. For females the within-trait intercept and slope correlation did not reach significance for internalizing (r = –.20, p = .383) nor externalizing (r = –.20, p = .059) behaviors. Males showed significant negative intercept–slope correlations for both internalizing (r = –.38, p = .009) and externalizing (r = –.23, p = .032) behaviors.

3. The model examining parent ratings initially produced a warning that the residual covariance matrix was nonpositive definite, due to a large residual correlation for the Year 7 internalizing and externalizing scores, which we resolved by imposing a boundary constraint (r < 1.0). This residual correlation for Year 7, as well as one for Year 15, sometimes exceeded 1.0 in other models with the parent ratings, so we bounded them to be below 1.0 throughout.

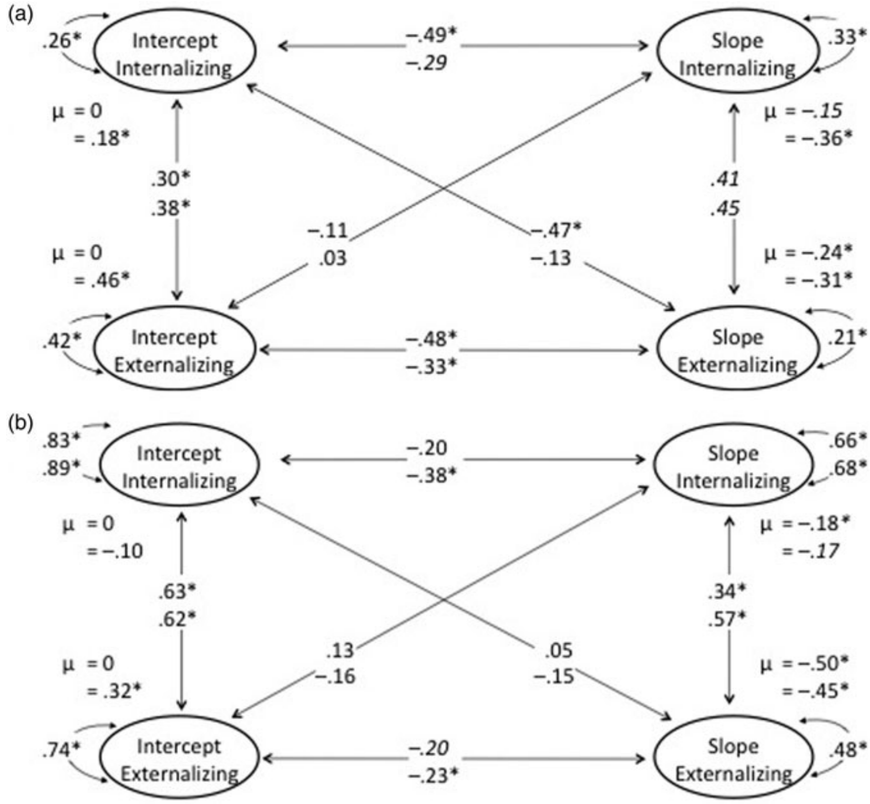

Figure 2. Bivariate growth models of internalizing and externalizing as rated by (a) teachers and (b) parents. Latent variable means and variances are unstandardized, whereas numbers on the double-headed arrows are correlations. Numbers on top are parameters for girls; those on the bottom are for boys. If only one number is present, that parameter was shown to be invariant and constrained across sex. $^*p < .05$; italic values indicate $p < .10$, as indicated by z tests formed from the ratio of the parameter divided by its standard error.

The cross-trait correlations between the intercepts and slopes of internalizing and externalizing behaviors indicate the degree of covariance. The internalizing and externalizing intercept factors were significantly correlated in both females ($r = .63$, $p < .001$) and males ($r = .62$, $p < .001$), as were the slope factors (females, $r = .34$, $p = .005$; males, $r = .57$, $p < .001$). We did not test whether these correlations were significantly different across sex, given the lack of invariance in the internalizing model.

Association between EFs and growth factors

To examine whether EFs could account for some of the covariance between internalizing and externalizing growth factors, we regressed the growth factors in each bivariate growth model on the three EF components, allowing separate regression parameters for males and females in the initial models. We then tested whether each association between the EFs and growth factors could be constrained to be equal across sex. Standardized betas for the best fitting models are provided in Table 1. Standard errors and p values for the standardized estimates are presented in the text.

Teacher ratings. Initially, we estimated a model in which we allowed the associations between teacher-rated intercepts and slopes and the three EF factors to vary across sex. This model fit well, χ^2 (696) = 721.83, $p = .241$, CFI = 0.990, RMSEA = 0.009.[4] The estimates for the teacher and parent ratings, estimated separately for males and females, are available in online-only supplementary Table S.5. Two paths were moderated by sex: the association between common EF and the intercept of externalizing, χ^2 difference (1) = 10.31, $p = .001$; and the association between the shifting-specific factor and intercept of internalizing, χ^2 difference (1) = 5.92, $p = .015$. All other associations could be constrained to be equal across sex, all χ^2 difference (1) > 2.36, $p > .123$.

The model shown in Table 1 constrains all but these two associations to be equal across sex at the unstandardized levels, as well as the residual variances for the slope factors.

4. This model produced a warning that the latent variable covariance matrix in males was not positive definite, noting that the problem involved the internalizing slope. We did not find any covariances that were out of bounds, so the warning likely reflects that all the variability in the slope was predicted by all other variables in the model, leading to a multiple correlation approaching 1 (i.e., a linear dependency among more than two latent variables). We did not receive a warning for the final P factor model with teacher ratings. Because the estimates for the regressions of the internalizing and externalizing growth factors were similar in this bivariate model to those we obtained in models including only internalizing or only externalizing (which did not produce any warnings), we report the results of the full model here.

Table 1. *Standardized regression coefficients (female/male) for growth factors regressed on executive functions (EFs)*

Growth Factors	Common EF	Updating Specific	Shifting Specific
Teacher ratings			
Intercept internalizing	–.26*/–.24*	.01/.01	.29*/–.17[a]
Intercept externalizing	.04/–.50*[a]	.19*/.19*	.14/.14
Intercept *r* predicted	–.01/.12*	.00/.00	.04/–.02
Slope internalizing	.09	–.15	–.16
Slope externalizing	–.23	.08	.40*
Slope *r* predicted	–.02	–.01	–.06
Parent ratings			
Intercept internalizing	–.17*	–.09	.01
Intercept externalizing	–.04	.00	.25*
Intercept *r* predicted	.01	.00	.00
Slope internalizing	.22†	.04	.10
Slope externalizing	–.14	–.01	–.10
Slope *r* predicted	–.03	.00	–.01

Note: Parent- and teacher-rating models were estimated separately. Unless noted, unstandardized parameters were constrained to be equal for males and females, but separate estimates are reported because standardized parameters differed slightly. Values in the *r predicted* rows describe the correlation between the internalizing and externalizing growth factors due to the common association with EF.
[a]Parameters for males and females were significantly different, so they were freely estimated.
†$p < .10$, as indicated by z tests formed from the ratio of the parameter divided by its standard error. *$p < .05$.

Because of the sex differences in these association between common EF and the intercept of externalizing, and shifting-specific and the intercept of internalizing, the residual variances of the intercepts were allowed to vary between sex, resulting in slightly different estimates for the regression betas of the intercepts on EFs when standardized.

This model fit the data well, χ^2 (704) = 738.65, $p = .177$, CFI = 0.986, RMSEA = 0.011, and was not significantly different from a model without these EF path constraints, χ^2 difference (10) = 15.34, $p = .120$. In line with our hypothesis, common EF was significantly associated with the internalizing intercept for both sexes (females, $\beta = –0.26$, $SE = 0.093$, $p = .005$; males, $\beta = –0.24$, $SE = 0.081$, $p = .003$). However, common EF significantly related to the externalizing intercept for males ($\beta = –0.50$, $SE = 0.106$, $p < .001$) but not females ($\beta = 0.04$, $SE = 0.097$, $p = .692$). Multiplying these paths for males indicated that common EF predicted a correlation of .12 ($SE = 0.053$, $p = .023$), which was 32% of the total .38 correlation. Because common EF did not predict the externalizing intercept for females, it did not predict a significant portion (predicted correlation = –.01, $SE = .025$, $p = .687$) of the .30 intercept correlation in females.

Updating-specific was significantly positively related to the externalizing intercept (both sexes, $\beta = 0.19$, $SE = 0.093$, $p = .038$). However, it was not related to the internalizing intercept (both sexes, $\beta = 0.01$, $SE = 0.102–0.111$, $p = .938$), so did not account for covariance.

There was a significant sex difference for shifting-specific and the intercept of internalizing, such that the association was significant in females ($\beta = 0.29$, $SE = 0.125$, $p = .018$) but not in males ($\beta = –0.17$, $SE = 0.112$, $p = .136$). Shifting-specific was not significantly related to the intercept of externalizing (females, $\beta = 0.14$, $SE = 0.085$, $p = .106$; males, $\beta = 0.14$, $SE = 0.084$, $p = .103$). Thus, shifting-specific did not account for a significant proportion of the intercept correlation.

Shifting-specific also uniquely predicted the externalizing slope (both sexes, $\beta = 0.40$, $SE = 0.176$, $p = .025$). Because the slope of internalizing behavior was not also related to this factor ($\beta = –0.16$, $SE = 0.133$, $p = .229$), shifting-specific did not explain a significant portion of the correlation between the internalizing and externalizing slope factors.

Parent ratings. We followed the same model building procedure with parent ratings. The initial model allowed for the associations between common EF and the growth factors to vary between sexes. The model fit well, χ^2 (680) = 751.17, $p = .030$, CFI = 0.995, RMSEA = 0.015. None of the associations between the EF factors and the growth factors differed by sex, all χ^2 difference (1) < 1.26, all $p > .261$.

The model shown in the lower portion of Table 1 constrained all the associations between the multiple components of EF and the growth factors to be equal across sex, χ^2 (692) = 786.26, $p = .007$, CFI = 0.993, RMSEA = 0.017, although it fit slightly worse than the unconstrained model, χ^2 difference (12) = 22.26, $p = .034$. Consistent with the model for teacher ratings, common EF was associated with the internalizing intercept ($\beta = –0.17$, $SE = 0.081$, $p = .036$) and shifting-specific was positively associated with the externalizing intercept ($\beta = 0.25$, $SE = 0.078$, $p = .002$). However, there was no evidence for common EF explaining internalizing with externalizing covariance in this model, as all associations seemed to be specific to each domain.

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

Association between EFs and a P factor with residual growth factors

To further investigate how common variance across internalizing and externalizing behaviors relates to EFs, we estimated a bifactor P factor in addition to our growth factors for teacher and parent ratings, as illustrated in Figure 3. We regressed all time points on the latent P factor in addition to the latent intercepts and slopes for internalizing and externalizing behaviors. The growth factors in the context of the P factor can be conceptualized as residuals, capturing variance specific to internalizing and externalizing trajectories. Because the P factor captures covariance in internalizing and externalizing behaviors across time, we constrained the cross-behavior intercept and slope correlations to be zero, but allowed the negative within-trait intercept and slope correlations to be free.

Teacher ratings. For teacher-rated data, the P factor loadings could be constrained to be equal across sex, χ^2 difference (18) = 24.72, $p = .133$; thus, in our final model, the P factor loadings were invariant. This full P factor model fit the data well, χ^2 (328) = 340.26, $p = .309$, CFI = 0.994, RMSEA = 0.010. All but one time point loaded significantly on the P factor (see online-only supplementary Table S.4 for standardized loadings), showing substantial shared influence in these traits across time.

We then estimated a model with the three EF factors correlated with the P factor and the specific intercepts and slopes, χ^2 (680) = 700.687, $p = .283$, CFI = 0.992, RMSEA = 0.008. Although single-degree of freedom difference tests indicated that the correlations between EFs and the behavior problem factors were not significantly different across sex, all χ^2 difference (1) < 1.24, $p > .266$, a model in which all correlations of the EFs with the P factor and growth factors were constrained across sex, χ^2 (695) = 819.37, $p < .001$, RMSEA = 0.020, CFI = 0.950, fit significantly worse than the model with sex differences in correlations, χ^2 difference (15) = 66.69, $p < .001$. That is, even though each correlation could be equated across sex when others were allowed to vary, constraining all associations at once led to a large decrement in fit because sex differences in correlations could not be absorbed by the P factor nor the specific factors.

Thus, we estimated a model allowing some sex differences based on the parameters we identified as differing across sex in the bivariate growth and EFs model presented in Table 1 (i.e., the relation between common EF and the externalizing intercept and between shifting-specific and the internalizing intercept). Specifically, we allowed sex differences in the correlations of common EF with the P factor and both intercepts, and the correlation of shifting-specific with the internalizing intercept. This model, presented in Table 2, χ^2 (691) = 707.58, $p = .323$, CFI = 0.993, RMSEA = 0.007, fit no worse than the model in which all the correla-

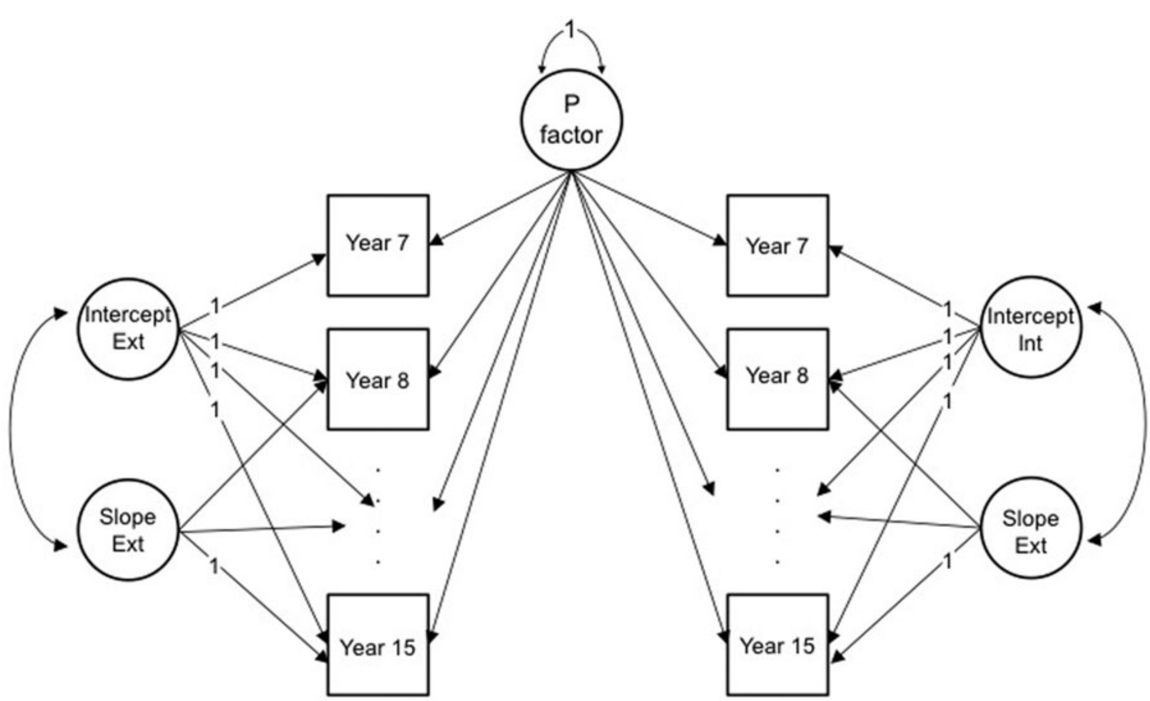

Figure 3. Specification of the psychopathology (P) factor model for teacher ratings. Scores for all time points (ellipses indicate additional time points not shown for clarity) were regressed on the P factor, with loadings constrained to be equal across sex. Internalizing and externalizing growth factors were specified following invariance testing reported in the Method section. Because the P factor accounts for cross-trait covariances, the intercept–intercept, slope–slope, and cross-trait intercept–slope correlations were fixed to zero; however, the negative within-trait intercept–slope correlations were estimated. Residual variances for each indicator and residual correlations were also estimated but are not shown.

Table 2. *Correlations (female/male) between executive functions (EFs), the psychopathology (P) factor, and growth factors*

	Executive Function Factor		
	Common EF	Updating Specific	Shifting Specific
Teacher ratings			
P factor	–.14/–.56*	.22	–.19
Intercept internalizing	–.05/–.17	–.10	.29†/.10
Intercept externalizing	.20/–.16	.06	.27†
Slope internalizing	.09	–.15	–.13
Slope externalizing	–.28†	.14	.30†
Parent ratings			
P factor	–.25/–.30	–.11/–.12	.34/.41
Intercept internalizing	.10	.01	–.48
Intercept externalizing	.25	.13	–.06
Slope internalizing	.28†	.06	.06
Slope externalizing	–.08	.02	–.21†

Note: All latent variable variances could be constrained across sex with no decrement in fit, except for the variance of the P factor in the parent-ratings model; although covariances with this factor were equated across sex, separate correlations are shown due to slight differences in the standardized estimates. For the teacher-ratings model, the presence of two correlation estimates indicates that the covariances were allowed to differ across sex. The intercept–intercept, slope–slope, and cross-trait intercept–slope correlations were fixed to zero, but the within-trait intercept–slope correlations were estimated. For teacher ratings the intercept–slope correlations were as follows: internalizing females, $r = –.55$, $p < .001$; internalizing males, $r = –.35$, $p = .046$; externalizing females, $r = –.55$, $p < .001$; externalizing males, $r = –.42$, $p = .007$. For parent ratings the intercept–slope correlations were as follows: internalizing females, $r = –.45$, $p = .004$; internalizing males, $r = –.34$, $p = .067$; externalizing females, $r = –.70$, $p < .001$; externalizing males, $r = –.35$, $p = .011$.
†$p < .10$, as indicated by z tests formed from the ratio of the parameter divided by its standard error. *$p < .05$.

tions with the EFs were allowed to differ by sex, χ^2 difference (11) = 8.32, $p = .684$, but it fit significantly better than the model in which all the correlations with the EFs were constrained to be equal across sex, χ^2 difference (4) = 38.36, $p < .001$.

In this model, we could constrain the common EF with P factor correlation to be equal across sex, χ^2 difference (1) = 1.33, $p = .250$, and the resulting correlation (r = –.36) was significant; however, when we did so, there emerged a significant positive association with the externalizing intercept in females ($r = .33$, $p = .033$), which was inconsistent with the null common EF–externalizing intercept correlation in the bivariate growth model. This positive correlation with the externalizing intercept in females was needed to offset the negative correlation with externalizing behaviors predicted by the negative correlation with the P factor. We determined that we could not constrain both the associations of common EF with the P factor and the externalizing intercept to be equal across sex without a significant decrement in fit compared to the model shown in Table 2, χ^2 difference (2) = 7.82, $p = .020$, because the correlation of common EF with the P factor was significantly different across sex, χ^2 difference (1) = 5.95, $p = .015$, in the context of a model in which the common EF with externalizing intercept was also equated across sex. Based on these tests, as well as the results of the bivariate growth model presented in Table 1, we interpret the final model shown in Table 2 as indicating a sex difference in the association between common EF and the P factor.

As shown in the top portion of Table 2, common EF significantly correlated with the P factor in males ($r = –.56$, $SE = 0.199$, $p = .005$), but not in females ($r = –.14$, $SE = 0.218$, $p = .516$). No other associations were significant. However, consistent with the bivariate growth model, we found marginally significant associations between the externalizing slope and common EF (both sexes, $r = –.28$, $SE = 0.148$, $p = .054$) and between shifting-specific and the externalizing-specific factors in both sexes (externalizing intercept, $r = .27$, $SE = 0.149$, $p = .068$; externalizing slope, $r = .30$, $SE = 0.162$, $p = .061$).

The sex difference in the correlation of the P factor with common EF is consistent with the sex differences we observed in the bivariate growth model, in that we only observed a relation between common EF and covariance in the internalizing and externalizing intercepts in males. However, the P factor model did not show the significant association of shifting-specific with females' internalizing intercept that we saw in the bivariate growth model. Part of this association may be accounted for by the P factor, such that splitting the internalizing intercept variance into that related to the P factor and that unique to internalizing may have reduced the correlations with shifting-specific.

Parent ratings. In the parent-rated data, a similar P factor model with invariant loadings across sex provided a significantly worse fit than one with loadings allowed to vary across sex, χ^2 difference (18) = 37.54, $p = .004$. However, a model with invariant loadings but P factor variances allowed to vary

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

across sex did not result in a decrement in fit, χ^2 difference (16) = 23.57, p = .099. Thus, our final model had sex-invariant loadings but sex differences in the P factor variance (with a larger variance in females), χ^2 (311) = 390.963, p = .001, RMSEA = 0.024, CFI = 0.994.

We then added the EFs to this model and allowed them to correlate with the P factor and the growth factors, χ^2 (663) = 712.21, p = .091, RMSEA = 0.013, CFI = 0.996. The associations between the EFs and behavior problem factors could all be constrained to be equal across sex, all χ^2 difference (1) < 2.85, p > .092. Although a model with all correlations with the EFs invariant across sex did show a small but significant decrement in fit, χ^2 difference (15) < 26.28, p = .035, we opted to go with this model because the differences were slight, as in the bivariate growth model for the parent ratings presented earlier. Thus, correlations for the sex-invariant model, χ^2 (678) = 751.59, p = .026, RMSEA = 0.015, CFI = 0.995, are shown in the lower portion of Table 2.

EFs were not significantly related to any of the behavioral problem factors (all ps > .061 for growth factors, all ps > .127 for P factor). The associations that were significant in the bivariate growth model (a negative association between common EF and the internalizing intercept and a positive association between shifting-specific and the externalizing intercept) seem to have been absorbed by the P factor, but did not reach significance when these intercept variances were split into that related to the P factor and that unique to internalizing and externalizing intercepts.

Consistency with latent class growth curve analyses (LGCAs)

Growth models assume that individual differences in the intercept and slope factors can be described with a continuous normal distribution. Other models, such as LGCAs, model trajectories as categorical latent variables: that is, variation in intercept and slope factors are due to mixtures of subpopulations with unique stability and change parameters within the total study population (Nagin, 1999; for details of these models in Mplus, see Jung & Wickrama, 2008). As such mixture models are a popular way of analyzing trajectories that can provide complementary and sometimes different information than standard growth models (particularly if growth factors are not normally distributed), we estimated an LGCA for each behavior problem and rater and examined how the identified classes scored on the latent EF factors. The results are presented in the online-only supplementary materials (see supplemental Latent Class Growth Curve Analyses section). Overall, the LGCA results showed similar patterns to the growth models (Table 1), but the effects were smaller and fewer were significant, likely due to lower power for the LGCAs (Bauer & Curran, 2003).

Discussion

We used a multicomponent model of EFs to decompose the covariance between growth factors for teacher- and parent-rated internalizing and externalizing behavior from ages 7 to 16 years. We found more associations of EF variables with teacher-rated behavior than parent-rated behavior. In boys, common EF was significantly and negatively related to both teacher-rated internalizing and externalizing stability (intercept) factors and accounted for a significant percentage (32%) of their covariance. In a model with a P factor, common EF significantly correlated with the P factor in males. In females, common EF only related to the teacher-rated internalizing intercept.

We also observed positive associations with the shifting-specific factor, such that better shifting-specific abilities were associated with less decrease in teacher-rated externalizing behaviors from ages 7 to 15 years in both sexes. Shifting-specific also showed a positive correlation with the teacher-rated internalizing intercept for females. The negative association between common EF and internalizing behavior, and the positive association between shifting-specific and externalizing behavior were also present for the intercepts in the model with parent ratings. These results suggest that common EF relates to internalizing across sex and raters, while shifting-specific relates to externalizing across sex and raters. Sex moderated the associations of externalizing and the P factor with common EF for teacher-rated behavior.

This study contributes to the literature on these problem behaviors in several ways. It is the first to examine how multiple components of EF, measured with a well-validated latent variable model, relate to internalizing, externalizing, and their covariance. In addition, it examined relations with latent growth models of these behavior problems to further understand how EFs relate to stability and change in these behaviors across childhood and adolescence. Moreover, it examined a P factor in the context of these growth factors, and related this P factor to EFs. Finally, it explored rater differences and sex differences in the relations of these behaviors to EFs, which have often been ignored in the context of covariance. We discuss each of these points in more detail below. Because we found few associations with the parent-ratings model, we focus our discussion on the teacher ratings. We also do not discuss the supplementary LGCA results, as our results show that a single class with continuously varying intercepts and slopes is likely the most powerful model for examining relations with EFs within these data.

Associations with multiple components of EF

While past research has found that specific tasks and EF factors are associated with variation common across psychiatric traits (e.g., Caspi et al., 2014; Martel et al., 2017), an innovation of this study is the incorporation of a latent variable model of multiple EF components to examine both common and broadband specific effects. Specifically, we assessed individual differences in common EF ability, as well as updating-specific and shifting-specific abilities, which have been shown to differentially relate to other cognitive and behavioral traits (Friedman & Miyake, 2017; Miyake & Friedman, 2012).

Consistent with our primary hypothesis, we found that common EF explained some of the internalizing and externalizing covariance, though only in boys. Specifically, common EF was significantly associated with males' stable variance (i.e., intercept factors) in both behavior problems, explaining 32% of their correlation (i.e., predicting a correlation of .12 out of a total correlation of .38). Furthermore, common EF correlated significantly with the P factor in males (r = –.56). The common EF factor is thought to tap the ability to actively maintain goals and use those goals to bias ongoing processing (Friedman & Miyake, 2017; Miyake & Friedman, 2012). In the context of internalizing and externalizing behaviors, such cognitive control may be needed to reduce activation of behaviors that would lead to negative outcomes, and/or to regulate emotions (Zelazo & Cunningham, 2007).

These associations are broadly consistent with prior work examining the P factor's relations to general EF abilities. Martel et al. (2017) found that a global EF factor significantly predicted children's P factor (β = –0.24) but not specific variance in fear, distress, or externalizing factors. Caspi et al. (2014), examining adults assessed longitudinally from ages 18 to 38 years, found that EF tests, as well as full-scale IQ and subfactors, were related to the P factor (rs = –.20 to .17), with few associations with specific factors in addition to the P factor. However, neither study reported testing for moderation by sex (although Caspi et al., 2014, controlled for sex, they did not report correlations with EF tasks separately for males and females).

Our results contrast somewhat with recent finding by Nigg et al. (2017). They reported that a general EF latent factor (based on Trail Making B, Wisconsin Card Sorting Test perseverations, Stroop interference, stop-signal reaction time, and reaction time variability) was related to an externalizing but not internalizing latent factor (based on diagnostic interviews) in adults. They interpreted this result to mean that general EF may be more related to particular disorders rather than common variance. However, they regressed EF on externalizing and internalizing factors, controlling for one another, rather than regressing both psychopathology factors on EF, as we did. This direction of regression means that P factor variance would be controlled for in each of their regression coefficients. Moreover, they did not examine sex moderation. Thus, inconsistencies with our results could be due to differences in models, ages examined, or methods for assessing psychopathology.

Few prior studies have examined specific components of EF in addition to common EF. Doing so in the current study revealed *positive* associations with the shifting-specific factor. In the bivariate growth model, shifting-specific related positively to change (slope) in externalizing behavior in both sexes, and also related to stable variance (intercept) in internalizing behaviors in females. The shifting-specific factor is thought to reflect the persistence of goal representations in the prefrontal cortex after they are no longer needed (Friedman & Miyake, 2017; Miyake & Friedman, 2012); better shifting-specific ability reflects less persistence. Our finding of positive associations with shifting-specific is consistent with our prior work suggesting that problem behaviors may be negatively related to common EF but positively related to shifting-specific (see Herd et al., 2014). This pattern may reflect a stability-flexibility trade-off, whereby poor goal maintenance and implementation (lower common EF) makes it easier it is to shift to a different goal (better shifting-specific). Per this interpretation, individuals with more problem behaviors may have less interference from no-longer-relevant goals when switching between tasks because they had poorer maintenance of those goals when they were relevant.

We also found a positive association between stable variance in externalizing behavior and updating-specific in both sexes. This effect was not predicted, and its directionality is difficult to interpret, in that it indicates that more externalizing behavior is related to better updating-specific performance (thought to reflect accuracy of gating the contents of working memory, or memory retrieval; Friedman & Miyake, 2017; Miyake & Friedman, 2012). This pattern has not emerged in prior studies with this model: though past work has shown positive relations between shifting-specific and behavior problems, the updating-specific component typically is unrelated to such behavior problems but is positively related to intelligence (Herd et al., 2014).

Deficits in common EF/inhibition (response inhibition is isomorphic with our common EF factor) have been broadly implicated in multiple psychiatric disorders (Snyder, Miyake, et al., 2015). Furthermore, the multiple-component EF model we used in this study has been related to several specific problem behaviors (subclinical) that load on these more general internalizing and externalizing factors, including attention problems (Friedman et al., 2007), and sleep problems (Friedman, Corley, Hewitt, & Wright, 2009), as well as a more general behavioral disinhibition factor (Young et al., 2009). The current study demonstrates that EFs also explain variance in more general internalizing and externalizing scores. Thus, the negative associations with individual problem behaviors found in prior studies may reflect these more general associations with internalizing and externalizing constructs.

EFs' relations to stability and change in problems

Another strength of this study is the use of latent growth models to examine stability and change in internalizing and externalizing behaviors. Incorporating longitudinal data allowed us to examine how common variance in different aspects of trajectories was related to later EFs.

With respect to these trajectories, our results were largely consistent with prior research. In the model for teacher-reported data, we found that externalizing behavior problems decreased across time, consistent with the results of Keiley et al. (2000) and Gilliom and Shaw (2004); however, in contrast to those same prior studies, we found that internalizing behavior problems also decreased across time. This inconsistency is likely due to our inclusion of older ages than these

https://doi.org/10.1017/S0954579417001602 Published online by Cambridge University Press

prior studies. Inspection of the individual time points shows an initial increase in the behavior problems, consistent with past literature. Furthermore, past results with the data used in this study have found that this decrease is specific to the CBCL/TRF, with depression and anxiety diagnosis increasing across adolescence in this same sample across time, while internalizing symptoms of the CBCL/TRF decrease; however, diagnosis and checklist measures remain significantly correlated across time despite these inconsistencies (Johnson, Whisman, Corely, Hewitt, & Rhee, 2012).

It is of the most importance for our analysis that the stability and change latent variables significantly covaried across internalizing and externalizing behaviors, consistent with past research (Lee & Bukowski, 2012). These results indicate that covariance across internalizing and externalizing behaviors extends beyond a single time point. Children show internalizing and externalizing covariance not only in their overall levels of these behaviors but also in the extent to which these behaviors change across time. A new contribution of the current study is that we related this common variance in both stability and change factors to EFs, and utilized the cross-trait variation across time to estimate a latent P factor.

We found that common EF related most strongly to common variance in the boys' intercept factors, which capture stability in problem behaviors. In terms of the slope factors, which capture change in problem behaviors, we found few associations with EFs. Thus, even though there is common variance in how these behaviors change across time, common variance in this change does not seem to be related to EFs, in contrast to the patterns seen with the intercepts. We found an analogous association in the P factor model, where common EF only related to the P factor in boys. Similar to a correlated intercept model, past research on the P factor shows high stability across time (Murray, Eisner, & Ribeaud, 2016; Snyder et al., 2017). Therefore, our common P factor model represents patterns of common liability across time that is analogous to our intercept factors. In terms of the association with EF, it gives converging evidence that common EF relates largely to stable common psychiatric variability across time.

Sex differences in relations of EFs to behavior problems

Past research on internalizing and externalizing has found mean differences between females and males in levels of internalizing and externalizing behaviors (Keiley et al., 2000). Although there were no sex differences in levels of EFs in late adolescence (Friedman et al., 2016), we estimated models for the sexes separately to consider how sex differences may play a role in the development of disorders and their links to EFs.

The sex difference that was consistent across the bivariate growth model and the P factor model was that common EF was related to covariance only in males. In the bivariate growth model, common EF related similarly to stable variance in internalizing for males and females, but was more strongly related to stable variance in externalizing in males. Similarly, in the P factor model, common EF only correlated with the P factor in males. While we do not know a specific reason for this sex difference, Zahn-Waxler, Shirtcliff, and Marceau (2008) argued in their review on developmental psychopathology and gender interactions that there are "heterotopic" patterns of developmental problem behaviors between sexes. These authors speculated that some cognitive factors are more associated in one sex due to the sexes experiencing different external social pressures and internal biological developmental patterns.

Of course, we have no data in the current study to support this mechanism, and it is possible that differences in internalizing and externalizing behavior (e.g., boys had higher mean levels for both behaviors in our study) predispose the sexes to different patterns of development of EF, whereby they develop different and less effective coping or regulation systems because of their underlying behavior problems.

We did detect some effects that did not vary by sex. Both the association between common EF and internalizing intercept, and the association between shifting-specific and the externalizing slope could be constrained across sex without a large determent in fit. Furthermore, it is important to note that the associations that could not be constrained across sex were only in the teacher-rated model, and it has yet to be shown if these sex differences would hold in diagnostic or self-report data.

Strengths, limitations, and future directions

A key limitation of this study is that, whereas internalizing and externalizing were measured longitudinally (ages 7–16 years), EFs were only measured at one age (age 17 years), and that time point was later than the problem behaviors. Thus, we could not quantify the extent to which EFs and problem behaviors related at the same time point, nor examine how the development of EFs paralleled the development of problem behaviors. Because this is a correlational study, we also could not disentangle the directionality of the associations. Hence, we do not make any claims regarding causality or direction of effect in the association between EFs and problem behaviors. However, that we observed associations between common EF and the intercepts of the growth models suggests that it is the stable variance in problems (in our parameterization, individual differences at age 7 and variance at later time points that is related to those initial levels) that is related to common EF. This pattern, combined with the high level of stability in EFs seen across this age range in other samples (Holmes, Kim-Spoon, & Deater-Deckard, 2016) and in the P factor in other samples (Snyder et al., 2017), raise the possibility that the associations we observed reflect stable variance in both behavioral problems and EFs.

In addition to the use of a multicomponent latent variable EF model and growth models, a strength of this study is assessment by multiple raters, because each type of rater has its own biases. Generally, teachers have been shown to be better raters of behavioral problems than parents (Lochman, 1995). While we were unable to test specific hypothesis about why EF is association with common psychopathology variance in teachers but not parents, past research has shown that parents, mostly mothers, show discrepancy with teacher reports and that these discrepancies may be unrelated to the child's problem behavior per se (Webster-Stratton, 1988). However, the low correlation between parent and teacher ratings may be only partly attributable to rater biases (Achenbach, McConaughy, & Howell, 1987), as these raters interact with the children in different environments (school vs. home). In this study, we found different results for parent- and teacher-rated behavior, suggesting that EFs relate to covariance and specificity of problem behaviors differently depending on the particular context in which behaviors are assessed. Ascertaining whether effects are consistent across raters may help constrain models of possible mechanisms that underlie these associations (e.g., EFs may be more related to the expression of internalizing and externalizing behavior under cognitively demanding conditions such as those found in classrooms).

The longitudinal twin study sample was not selected for any behavior problems, so the use of a checklist measure of problems, rather than clinical interview, was an appropriate method to assess continuous variation in these behaviors. However, there is some reason to believe these results might generalize to samples that include more extreme psychiatric variation. For example, such childhood problem behaviors relate significantly to diagnosis (Bilenberg, 1999) and are a robust predictor of adult psychiatric traits (Hofstra, van der Ende, & Verhurlst, 2000). As described earlier, the results of Caspi et al. (2014) and Martel et al. (2017) suggest that we might observe similar patterns for latent factors based on clinical diagnoses and for covariance at later ages, including adulthood.

Although we use data from twins, we focused here on the phenotypic associations, as a first step to examine how EFs predict common variance in these problems. Future research could analyze these associations at the genetic level.

Finally, if twins have unique patterns of behavior, this study would not generalize to the rest of the population. However, previous studies have found that twins tend to be representative of the general population with respect to common psychiatric symptoms (Kendler, Martin, Heath, & Eaves, 1995); thus, these results should generalize to nontwin samples.

Alternative predictors of internalizing–externalizing covariance

Although shared variation with EFs explained 32% of the covariance in the stable variance in boys' internalizing and externalizing behaviors, the remaining 68% of this covariance was unexplained, as was covariance of the change in these behaviors. Therefore, traits beyond those assessed in our model of EF must explain some of the covariance in internalizing and externalizing behavior.

Some candidates to explain this remaining covariance are personality and emotion-related processes, particularly neuroticism/negative emotionality. This trait is considered to measure features that predispose to negative emotional states, higher emotional arousal, and more maladaptive coping mechanisms that have been considered characteristic of internalizing and externalizing disorders more broadly (e.g., Rhee et al., 2015). Furthermore, neuroticism/negative emotionality has been shown to explain covariance in internalizing and externalizing behaviors at both the phenotypic and genetic levels (e.g., Hink et al., 2013; Tackett et al., 2013). While there is overlap between EFs and neuroticism, the two traits are not isomorphic, and do retain significant amounts of unique variation (Fleming, Heintzelman, & Bartholow, 2015). Furthermore, self-reported cognitive control and neuroticism have been shown to explain independent portions of the covariance between depression and alcoholism, representative disorders of internalizing and externalizing (Ellingson, Richmond-Rakerd, & Slutske, 2015).

Conclusions

We found that common EF predicted the covariance in teacher-rated internalizing and externalizing behaviors that was stable from ages 7 to 15 years in boys but not in girls, whereas a shifting-specific factor was related primarily to externalizing behaviors (across sex). Future research should elucidate why boys and girls showed these different patterns, what mechanisms or pathways may mediate these relations, and what other factors explain covariance between internalizing and externalizing behavior.

Supplementary Material

To view the supplementary material for this article, please visit https://doi.org/10.1017/S0954579417001602.

References

Achenbach, T. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. (1991b). *Manual for the Teacher Report Form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232. doi:10.1037/0033-2909.101.2.213

Arseneault, L., Moffitt, T. E., Caspi, A., Taylor, A., Rijsdijk, F. V, Jaffee, S. R., . . . Measelle, J. R. (2003). Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *Journal of Child Psychology and Psychiatry*, *44*, 832–848. doi:10.1111/1469-7610.00168

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363. doi:10.1037/1082-989X.8.3.338

Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: Standardization and validation in Danish population based and clinically based samples. *Acta Psychiatrica Scandinavica*, *100*, 2–52. doi:10.1111/j.1600-0447.1999.tb10703.x

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The P factor. *Clinical Psychological Science*, *2*, 119–137. doi:10.1177/2167702613497473

Cerdá, M., Sagdeo, A., & Galea, S. (2008). Comorbid forms of psychopathology: Key patterns and future research directions. *Epidemiologic Reviews*, *30*, 155–177. doi:10.1093/epirev/mxn003

Cohen, N. J., Gotlieb, H., Kershner, J., & Wehrspann, W. (1985). Concurrent validity of the internalizing and externalizing profile patterns of the Achenbach Child Behavior Checklist. *Journal of Consulting and Clinical Psychology*, *53*, 724–728. doi:10.1037/0022-006X.53.5.724

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*, 584–594. doi:10.1037/0022-006X.55.4.584

Cosgrove, V. E., Rhee, S. H., Gelhorn, H. L., Boeldt, D., Corley, R. C., Ehringer, M. A., & Hewitt, J. K. (2011). Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *Journal of Abnormal Child Psychology*, *39*, 109–123. doi:10.1007/s10802-010-9444-8

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research*, *7*, 659–669. doi:10.1375/twin.7.6.659

Derks, E. M., Hudziak, J. J., Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics*, *34*, 571–583. doi:10.1007/s10519-004-5585-2

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. doi:10.1146/annurev-psych-113011-143750

Ellingson, J. M., Richmond-Rakerd, L. S., & Slutske, W. S. (2015). Brief report: Cognitive control helps explain comorbidity between alcohol use disorder and internalizing disorders. *Journal of Studies on Alcohol and Drugs*, *76*, 89–94. doi:10.15288/jsad.2015.76.89

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2015). Specifying associations between conscientiousness and executive functioning: Mental set shifting, not prepotent response inhibition or working memory updating. *Journal of Personality*, *84*, 348–360. doi:10.1111/jopy.12163

Friedman, N. P., Corley, R. P., Hewitt, J. K., & Wright, K. P. (2009). Individual differences in childhood sleep problems predict later cognitive executive control. *Sleep*, *32*, 323–333.

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science*, *18*, 893–900. doi:10.1111/j.1467-9280.2007.01997.x

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. doi:10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*, 326–340. doi:10.1037/dev0000075

Friedman, N. P., Miyake, A., Young, S. E., Defries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. doi:10.1037/0096-3445.137.2.201

Gilliom, M., & Shaw, D. S. (2004). Codevelopment of externalizing and internalizing problems in early childhood. *Development and Psychopathology*, *16*, 313–333. doi:10.1017/S0954579404044530

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: Advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, *23*, 41–57. doi:10.1002/mpr.1410

Herd, S. A., O'Reilly, R. C., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia*, *62*, 375–389. doi:10.1016/j.neuropsychologia.2014.04.014

Hink, L. K., Rhee, S. H., Corley, R. P., Cosgrove, V. E., Hewitt, J. K., Schulz-Heik, R. J., & Waldman, I. D. (2013). Personality dimensions as common and broadband-specific features for internalizing and externalizing disorders. *Journal of Abnormal Child Psychology*, *41*, 939–957. doi:10.1007/s10802-013-9730-3

Hinshaw, S. P. (2002). Process, mechanism, and explanation related to externalizing behavior in developmental psychopathology. *Journal of Abnormal Child Psychology*, *30*, 431–446. doi:10.1023/A:1019808712868

Hofstra, M. B., van der Ende, J., & Verhulst, F. C. (2000). Continuity and change of psychopathology from childhood into adulthood: A 14-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 850–858. doi:10.1097/00004583-200007000-00013

Holmes, C. J., Kim-Spoon, J., & Deater-Deckard, K. (2016). Linking executive function and peer problems from early childhood through middle adolescence. *Journal of Abnormal Child Psychology*, *44*, 31–42.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453. doi:10.1037/1082-989X.3.4.424

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, *108*, 663–676. doi:10.1016/j.jecp.2010.06.005

Johnson, D. P., Whisman, M. A., Corley, R. P., & Hewitt, J. K., & Rhee, S. H. (2012). Association between depressive symptoms and negative dependent life events from late childhood to adolescence. *Journal of Abnormal Child Psychology*, *40*, 1385–1400. doi:10.1007/s10802-012-9642-7

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302–317. doi:10.1111/j.1751-9004.2007.00054.x

Kazdin, A. E., & Kagan, J. (1994). Models of dysfunction in developmental psychopathology. *Clinical Psychology: Science and Practice*, *1*, 35–52. doi:10.1111/j.1468-2850.1994.tb00005.x

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, *28*, 161–179. doi:10.1023/A:1005122814723

Keiley, M. K., Lofthouse, N., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2003). Differential risks of covarying and pure components in mother and teacher reports of externalizing and internalizing behavior across ages 5 to 14. *Journal of Abnormal Child Psychology*, *31*, 267–283. doi:10.1023/A:1023277413027

Kendler, K. S., Martin, N. G., Heath, A. C., & Eaves, L. J. (1995). Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different? *American Journal of Medical Genetics*, *60*, 588–591. doi:10.1002/ajmg.1320600622

Kim, J., & Cicchetti, D. (2006). Longitudinal trajectories of self-system processes and depressive symptoms among maltreated and nonmaltreated children. *Child Development*, *77*, 624–639. doi:10.1111/j.1467-8624.2006.00894.x

Kroes, M., Kalff, A. C., Steyaert, J., Kessels, G. H., Feron, F. J. M., Hendriksen, J. G. M., . . . Vles, J. S. H. (2002). A longitudinal community study: Do psychosocial risk factors and child behavior checklist scores at 5 years of age predict psychiatric diagnoses at a later age? *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 955–963. doi:10.1097/00004583-200208000-00014

Krueger, R. F., & Tackett, J. L. (2003). Personality and psychopathology: Working toward the bigger picture. *Personality, Personality Disorder, and Psychopathology*, *17*, 109–128. doi:10.1521/pedi.17.2.109.23986

Laceulle, O. M., Vollebergh, W. A. M., & Ormel, J. (2015). The structure of psychopathology in adolescence: Replication of a general psychopathology factor in the TRAILS study. *Clinical Psychological Science*, *3*, 850–860. doi:10.1177/2167702614560750

Lee, E. J., & Bukowski, W. M. (2012). Co-development of internalizing and externalizing problem behaviors: Causal direction and common vulnerability. *Journal of Adolescence*, *35*, 713–729. doi:10.1016/j.adolescence.2011.10.008

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology*, *31*, 285–291. doi:10.1023/A:1023229529866

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

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335–361. doi:10.1177/014662168901300402

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, 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. *Journal of Abnormal Psychology*, *126*, 137–148. doi:10.1037/abn0000205

McGrath, L. M., Braaten, E. B., Doty, N. D., Willoughby, B. L., Wilson, H. K., O'Donnell, E. H., . . . Doyle, A. E. (2016). Extending the "cross-disorder" relevance of executive functions to dimensional neuropsychiatric traits in youth. *Journal of Child Psychology and Psychiatry*, *57*, 462–471. doi:10.1111/jcpp.12463

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. doi:10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review*, *20*, 113–136. doi:10.1016/S0272-7358(98)00096-8

Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The development of the general factor of psychopathology "P factor" through childhood and adolescence. *Journal of Abnormal Child Psychology*, *44*, 1573–1586. doi:10.1007/s10802-016-0132-1

Muthén, L. K., & Muthén, B. O. (1998–2014). *Mplus user's guide* (7th ed.). Los Angeles: Author.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157.

Nigg, J. T., Jester, J. M., Stavro, G. M., Ip, K. I., Puttler, L. I., & Zucker, R. A. (2017). Specificity of executive functioning and processing speed problems in common psychopathology. *Neuropsychology*, *31*, 448–466. doi:10.1037/neu0000343

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A heuristic for developing transdiagnostic models of psychopathology: Explaining multifinality and divergent trajectories. *Perspectives on Psychological Science*, *6*, 589–609. doi:10.1177/1745691611419672

Oglivie, J. M., Steward, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology*, *49*, 1063–1107. doi:10.1111/j.1745-9125.2011.00252.x

Rhea, S. A., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2013). Colorado twin registry: An update. *Twin Research and Human Genetics*, *16*, 351–357. doi:10.1017/thg.2012.93

Rhee, S. H., Lahey, B. B., & Waldman, I. D. (2015). Comorbidity among dimensions of childhood psychopathology: Converging evidence from behavior genetics. *Child Development Perspectives*, *9*, 26–31. doi:10.1111/cdep.12102

Riggs, N. R., Blair, C. B., & Greenberg, M. T. (2004). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology*, *9*, 267–276. doi:10.1076/chin.9.4.267.23513

Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: Unifying issues and research strategies. *Journal of Child Psychology and Psychiatry*, *44*, 1092–1115. doi:10.1111/1469-7610.00194

Sergeant, J. A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behavioural Brain Research*, *130*, 3–28. doi:10.1016/S0166-4328(01)00430-2

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*, 81–132. doi:10.1037/a0028727

Snyder, H. R., Kaiser, R. H., Warren, S. L., & Heller, W. (2015). Obsessive-compulsive disorder is associated with broad impairments in executive function: A meta-analysis. *Clinical Psychological Science*, *3*, 301–330. doi:10.1177/2167702614534210

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*, 328. doi:10.3389/fpsyg.2015.00328

Snyder, H. R., Young, J. F., & Hankin, B. L. (2017). Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clinical Psychological Science*, *5*, 98–110. doi:10.1177/2167702616651076

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. doi:10.1037/a0034151

Webster-Stratton, C. (1988). Mothers' and fathers' perceptions of child deviance: Roles of parent and child behaviors and parent adjustment. *Journal of Consulting and Clinical Psychology*, *56*, 909–915. doi:10.1037/0022-006X.56.6.909

Weiss, B., Susser, K., & Catron, T. (1998). Common and specific features of childhood psychopathology. *Journal of Abnormal Psychology*, *107*, 118–127. doi:10.1037/0021-843X.107.1.118

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*, 117–130. doi:10.1037/a0014657

Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology*, *4*, 275–303. doi:10.1146/annurev.clinpsy.3.022806.091358

Zelazo, P. D., & Cunningham, W. A. (2007). Executive function: Mechanisms underlying emotion regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 135–158). New York: Guilford Press.