Supplemental Materials for
Do Executive Functions Explain the Covariance Between Internalizing and Externalizing Behaviors?

Supplemental Table 1
Descriptive Statistics for Problem Behaviors by Rater and Sex of Child

Year	Age in Years (SD)	Mean Internalizing (SD)	Mean Externalizing (SD)	*N*
Teacher Ratings				
Female children				
Year 7	7.39 (.36)	5.27 (5.76)	3.91 (7.11)	289
Year 8	8.38 (.34)	4.33 (4.78)	3.31 (5.66)	270
Year 9	9.40 (.38)	4.87 (5.53)	3.97 (7.35)	261
Year10	9.93 (.38)	4.82 (5.77)	3.20 (5.56)	253
Year 11	11.38 (.38)	4.04 (4.58)	2.88 (5.67)	249
Year 12	12.40 (.37)	4.32 (5.11)	2.43 (4.43)	234
Year 13	12.87 (.44)	4.48 (5.34)	2.27 (5.36)	203
Year 14	13.90 (.40)	4.12 (4.37)	2.32 (4.37)	188
Year 15[a]	14.81 (.40)	3.52 (5.24)	2.37 (4.94)	136
Male children				
Year 7	7.50 (.39)	5.42 (5.89)	7.03 (9.50)	285
Year 8	8.47 (.37)	6.00 (6.42)	7.18 (9.03)	263
Year 9	9.49 (.39)	5.47 (6.19)	6.51 (8.58)	248
Year10	9.99 (.41)	5.53 (5.70)	6.00 (8.21)	258
Year 11	11.41 (.36)	5.17 (6.28)	6.32 (8.52)	252
Year 12	12.47 (.38)	4.19 (5.49)	5.23 (8.39)	204
Year 13	12.98 (.45)	4.70 (5.72)	5.06 (8.15)	172
Year 14	13.97 (.43)	4.36 (5.79)	5.76 (8.57)	167
Year 15[a]	14.90 (.35)	3.93 (5.02)	3.36 (5.40)	120
Parent Ratings				
Female children				
Year 7	7.43 (.36)	4.90 (4.50)	6.63 (5.72)	319
Year 9	9.40 (.38)	5.13 (5.19)	6.43 (6.24)	327
Year10	9.93 (.38)	4.92 (4.85)	5.78 (5.77)	299
Year 11	11.38 (.38)	4.39 (4.96)	4.65 (5.13)	234
Year 12	12.40 (.37)	5.71 (6.12)	5.94 (6.60)	340
Year 13	12.87 (.44)	4.90 (5.51)	5.23 (6.16)	273
Year 14	13.90 (.40)	5.34 (6.03)	5.10 (6.86)	260
Year 15[a]	14.81 (.40)	4.34 (5.41)	3.80 (4.83)	186
Year 16	16.59 (.83)	6.10 (6.51)	5.46 (6.82)	322
Male children				
Year 7	7.43 (.36)	4.64 (4.61)	8.75 (7.08)	308
Year 9	9.49 (.39)	4.92 (5.11)	8.25 (7.07)	311

Year10	9.99 (.41)	4.85 (4.96)	7.47 (6.81)	279
Year 11	11.41 (.36)	4.92 (5.32)	7.81 (7.41)	258
Year 12	12.47 (.38)	5.09 (4.86)	7.59 (6.88)	312
Year 13	12.98 (.45)	4.85 (5.49)	7.27 (7.39)	233
Year 14	13.97 (.43)	4.52 (4.85)	6.90 (6.90)	223
Year 15[a]	14.90 (.35)	3.89 (5.24)	5.84 (7.02)	165
Year 16	16.57 (.75)	4.99 (5.68)	7.56 (8.56)	313

[a]For twins whose 16th birthdays were within 4 months of when the age 15 assessment would have been completed, the age 15 assessment was skipped, resulting in a smaller *N* for that year.

Supplemental Table 2

Descriptive Statistics for Executive Function Tasks

EF Task	N	Mean	SD	Min	Max	Skewness	Kurtosis	Reliability
Antisaccade[a]	779	1.04	0.20	0.47	1.57	–0.12	–0.26	.89[b]
Stop-signal	741	282 ms	63	151	489	1.13	1.51	.75[b]
Stroop	759	214 ms	90	0	488	0.59	0.19	.91[b]
Keep track[a]	774	0.94	0.18	0.38	1.49	0.31	0.56	.65[c]
Letter memory[a]	785	1.09	0.25	0.38	1.57	0.29	–0.20	.62[c]
Spatial 2-back[a]	777	1.17	0.17	0.65	1.57	–0.93	1.65	.90[c]
Number–letter	776	331 ms	183	–14	923	1.04	1.12	.86[b]
Color–shape	768	331 ms	189	–196	916	0.76	0.75	.85[b]
Category-switch	766	333 ms	181	–34	899	0.98	0.92	.83[b]

Note. Table reproduced from Friedman et al. (2016), with permission.
Min = minimum; Max = maximum.
[a]Accuracy scores were arcsine transformed.
[b]Internal reliability was calculated by adjusting split-half or part1–part2 correlations with the Spearman–Brown prophecy formula.
[c]Internal reliability was calculated using Cronbach's alpha.

Supplemental Table 3

Unstandardized Loadings for Slope Factors in Bivariate Models

	Teacher Ratings[a]		Parent Ratings		
Year[b]	Externalizing	Internalizing	Externalizing[a]	Internalizing Female	Internalizing Male
Year 7	0	0	0	0	0
Year 8	–0.13	–0.11	--	--	--
Year 9	–0.04	0.24	0.39*	0.80*	0.25*
Year 10	0.09	0.15	0.55*	0.78*	0.30*
Year 11	0.19	0.52*	0.70*	0.93*	0.51*
Year 12	0.54*	0.78*	0.66*	0.88*	0.67*
Year 13	0.90*	0.46*	0.84*	1.05*	0.58*
Year 14	0.71*	0.95*	1.06*	1.25*	0.90*
Year 15	1	1	1.44*	1.29*	1.07*
Year 16	--	--	1	1	1

Note. Bivariate growth models did not include EFs. -- indicates that data were not available for that year for that rater.

[a]Models for the teacher ratings and for the externalizing parent-ratings were sex invariant.

[b]Loadings for the first and last time points were fixed to 0 and 1, respectively, to identify the Slope factors. With this parameterization, scores on the Slope factor can be interpreted as the total change across the time points examined, and each estimated loading represents the proportion of that total change at that age.

*$p<.05$, as indicated by z-tests formed from the ratio of the parameter divided by its standard error.

Supplemental Table 4

Standardized Loadings for Internalizing and Externalizing Scores on P Factor

	Teacher Ratings		Parent Ratings	
Year	Externalizing	Internalizing	Externalizing	Internalizing
Year 7	0.27*	0.19*	0.62*	0.58*
Year 8	0.33*	0.20*	--	--
Year 9	0.73*	0.43*	0.76*	0.72*
Year 10	0.52*	0.33*	0.77*	0.65*
Year 11	0.40*	0.44*	0.81*	0.74*
Year 12	0.31*	0.23*	0.79*	0.68*
Year 13	0.37*	0.14	0.74*	0.66*
Year 14	0.35*	0.26*	0.78*	0.69*
Year 15	0.33*	0.23*	0.70*	0.62*
Year 16	--	--	0.70*	0.65*

Note. Loadings were sex-invariant so constrained across sex. -- indicates that data were not available for that year for that rater.

*p<.05, as indicated by z-tests formed from the ratio of the parameter divided by its standard error.

Supplemental Table 5

Standardized Regression Coefficients (Female/Male) for Growth Factors Regressed on EFs in Models Without Sex Invariance for Regression Paths

	EF Factor		
Growth Factors	Common EF	Updating-Specific	Shifting-Specific
Teacher Ratings			
Intercept internalizing	*–0.20*/–.27*	0.14/–0.22	0.24*/–0.18
Intercept externalizing	–0.03/–.42*	0.30*/0.09	0.24*/0.03
Intercept *r* predicted	0.01/*0.11*	0.04/–0.02	0.06/–0.01
Slope internalizing	0.20/–0.08	–0.17/–0.07	*–0.30*/0.04
Slope externalizing	–0.13/*–.29*	–0.12/0.22	0.20/0.52*
Slope *r* predicted	–0.03/0.02	0.02/–0.02	–0.06/0.02
Parent Ratings			
Intercept internalizing	0.11/–.24*	–0.15/–0.01	–0.06/–0.01
Intercept externalizing	0.03/–0.06	0.10/–0.09	0.27*/0.13
Intercept *r* predicted	0.00/0.01	–0.02/0.00	–0.02/0.00
Slope internalizing	–0.06/0.10	0.09/–0.10	0.14/0.16
Slope externalizing	–0.08/–0.33*	–0.05/0.00	–0.16/0.15
Slope *r* predicted	0.00/–0.03	-0.01/0.00	–0.02/0.02

Note. Standardized path coefficients for the growth factors regressed on the executive function (EF) factors. Values in the "*r* predicted" rows describe the correlation between the internalizing and externalizing growth factors due to the common association with EF. Parent- and teacher-rating models were estimated separately.

*****$p<.05$, and italics font indicates $p<.10$, as indicated by *z*-tests formed from the ratio of the parameter divided by its standard error.

Supplemental Latent Class Growth Curve Analyses

Growth models assume that individual differences in the Intercept and Slope factors can be described with a continuous normal distribution. Other models, such as latent class growth curve analysis (LCGA), model trajectories as categorical latent variables: i.e., 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; see Jung & Wickrama, 2008, for details of these models in Mplus). 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 in fact 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.

First, to identify the number of classes for each behavior problem, we estimated a series of 1- through 4-class LGCA models for each behavior and each rater, without EFs in the model. We included sex as a known class, which splits each estimated class into two groups (by sex) that are allowed to differ on growth factor means. We used the same growth model (freed Slope loadings for all but the first and last time points), which was constrained to be invariant across classes. Variances and covariances of the growth factors were fixed to zero within classes, so differences in the Intercept and Slope means across the classes accounted for the total variation and covariation of the growth factors. We used the Bayesian Information Criterion (BIC) to decide on the number of classes (where the best model is the one with the lowest BIC; Jung & Wickrama, 2008), but also considered the size of the smallest class, as solutions with very few individuals (e.g., < 5% of the sample) in a class may not be useful. As shown in supplemental Table 6, for both problem types and raters, a 3-class solution seemed to provide the best fit, and these solutions showed acceptable separations of the classes (entropies = .684 to .777).

Next, we added the EF latent variables as distal outcomes to each 3-class LGCA. Trajectories for the 3-class teacher-rated and parent-rated problems from these models are available in supplemental Figures 1 and 2; these trajectories were very similar to those obtained without EFs in the model. Generally, the classes were distinguished by their Intercepts, with a few Slope differences, in line with some of the past literature on parent ratings across childhood and adolescence (Bongers, Koot, Van der Ende, Verhulst, 2004).

In these models, the EF model was strictly invariant across classes, with latent EF variances of 1.0. Thus, only the EF latent means were allowed to differ across class and sex (in addition to the growth factor means). By default, the latent EF means for the last group (Class 3 for males) were constrained to zero, so all other means represent the difference from that class in standard deviation units; this reference class corresponded to the male class that had the lowest level of problems in each model. As shown in supplemental Table 7 and supplemental Figures 1-2, there were some differences in Common EF across the classes based on teacher ratings, but not parent ratings. Specifically, boys in the class with the highest teacher-rated internalizing problems (7.0% of the sample) showed significantly lower Common EF ($\mu = -1.41$, $p=.004$) than boys in the class with the lowest problems (10.3% of the sample). Moreover, boys in the class with the highest teacher-rated externalizing problems (8.6% of the sample, $\mu = -1.22$, $p=.001$), as well as the boys in the class with an intermediate level of externalizing problems (23.2% of the sample; $\mu = -0.73$, $p=.010$) both showed significantly lower Common EF than boys in the class with the lowest problems (17.3% of the sample). Girls in all classes of teacher-rated externalizing behaviors showed similar mean differences from the reference class of boys, with the difference for the largest female class reaching significance (Class 2, consisting of 24.0% of the sample, μ= –0.47, p=.028). No other EF differences were significant in the teacher- and parent-ratings models, although there were non-significant trends for higher levels of problems to be associated with lower Common EF in most of the models.

Overall, the patterns seen in the LGCAs echo those seen in the full growth models (main text Table 1 and supplemental Table 5), but the effects were smaller and fewer were significant. This difference is likely due to the fact that these LGCAs capture individual differences in Intercepts and Slopes with 3 homogenous groups, akin to splitting a continuous distribution into high, medium, and low groups based on *z*-scores. When the estimated classes capture distinct patterns that cannot be described by correlated continuous growth factors (e.g., when the growth factors interact to predict outcomes), the LGCA can reveal new patterns; however if the LGCA simply artificially segments continuous variables, it reduces power (Bauer & Curran, 2003). Given that the classes we observed primarily reflected Intercept differences, the assumption of normally distributed growth factors from the models presented earlier is likely valid. Thus, the results presented in Tables 1 and 2 provide the most powerful test of our hypothesis that covariation in problem behaviors would be related to Common EF.

References Used in Supplement

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. http://doi.org/10.1037/1082-989X.8.3.338

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. http://doi.org/10.1111/j.1751-9004.2007.00054.x

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

Supplemental Table 6

Fit Statistics for Latent Class Growth Models With 1 to 4 Classes, Without EFs in the Models

Model	BIC	Entropy	Smallest Class	Largest Class
Teacher ratings				
Externalizing				
1 Class	11606.643	1.000	.497 (males)	.503 (females)
2 Classes	10835.457	0.845	0.224	0.276
3 Classes	**10759.272**	**0.777**	**0.081**	**0.238**
4 Classes	10754.052	0.724	0.039	0.182
Internalizing				
1 Class	11872.963	1.000	.497 (males)	.503 (females)
2 Classes	11639.889	0.791	0.129	0.372
3 Classes	**11592.045**	**0.736**	**0.062**	**0.325**
4 Classes	11596.709	0.647	0.063	0.242
Parent ratings				
Externalizing				
1 Class	14245.184	1.000	.494 (males)	.506 (females)
2 Classes	11301.714	0.822	0.207	0.294
3 Classes	**11255.901**	**0.743**	**0.078**	**0.239**
4 Classes	11275.181	0.689	0.037	0.206
Internalizing				
1 Class	14076.157	1.000	.494 (males)	.506 (females)
2 Classes	12030.520	0.746	0.154	0.349
3 Classes	**12007.089**	**0.684**	**0.056**	**0.291**
4 Classes	--	--	--	--

Note. Separate models were run for each behavior and each rater. All models were run using sex as a known class, which splits each class into two (males and females); thus the 1-class solution actually has 2 separate classes for males and females. The class sizes shown are based on the smallest and largest classes when classes are split by sex (proportions based on the estimated model). The latent growth model included an Intercept factor (unstandardized loadings of 1 for all time points) and a Slope factor (loading of zero for the first time point, 1 for the last time point, and free loadings for the remaining time points). Unstandardized loadings, residual variances, and thresholds were constrained to be equal across classes and sex, and factor variances were not allowed within class or sex. Thus, only the means of the latent growth factors differed across classes and sex, with the Intercept mean for the last class in males constrained to zero as a reference. The 4-class model for the parent-rated internalizing behaviors did not converge on an acceptable solution due to empty cells in the cross-tabs for the bivariate relations between ages of the ordinal behavior problem variables. BIC = Bayesian Information Criterion; lower numbers indicate a better fit, considering model complexity. In addition to BIC, we also considered the size of the smallest class in determining the best solution (indicated in bold face type), as classes smaller than 5% of the sample may not be useful. Thus, the 3-class solution was the best for all 4 models.