*Empirical Article*

Clinical Psychological Science
1–21

DOI: 10.1177/21677026251357589
www.psychologicalscience.org/CPS

# Longitudinal Clustering of Psychopathology Across Childhood and Adolescence: An Approach Toward Developmentally Based Classification

**Connor Lawhead[1], Jamilah Silver[1], Thomas M. Olino[2], Loïc Labache[3], Swanie Juhng[4], H. Andrew Schwartz[4], and Daniel N. Klein[1]**

[1]Department of Psychology, Stony Brook University; [2]Department of Psychology, Temple University; [3]Department of Psychiatry, Brain Health Institute, Rutgers University; and [4]Department of Computer Science, Stony Brook University

## Abstract

Current classification systems of psychopathology focus on cross-sectional symptoms rather than continuity, discontinuity, and comorbidity across development. Here, a community sample of 600 youths was assessed every 3 years from early childhood through late adolescence using semistructured diagnostic interviews. We used longitudinal *k*-means clustering of joint-diagnostic trajectories to identify six distinct clusters (healthy, childhood anxiety, childhood/adolescent attention-deficit/hyperactivity disorder, adolescent depression/anxiety, adolescent depression/substance use, and early childhood disruptive behavior). Comparing psychopathology clusters with the healthy cluster on age-3 predictors (parental education and psychopathology, early environment, temperament, cognitive and social functioning) and age-18 functional outcomes, we found that the clusters captured developmental patterning of psychopathology not apparent in cross-sectional nosology. The study serves as a proof of principle in applying a longitudinal clustering approach to common mental disorders, affording a rich perspective on the unfolding of sequential comorbidity and heterotypic continuity and identifying transdiagnostic subgroups with meaningful clinical, family, and temperamental correlates.

## Keywords

psychopathology, childhood, adolescence, clustering, classification, development

Received 11/21/24; Revision accepted 6/17/25

Existing classification systems of psychopathology vary in their consideration of development, but they are largely cross-sectional, emphasizing current psychopathology. However, some researchers have called for greater emphasis on development and course to better understand syndromes or symptom dimensions as they unfold and interact with one another (Oldehinkel & Ormel, 2023). An explicit focus on disorder continuities and discontinuities could shed light on distinct patterns of multifinality and equifinality, which would inform classification of psychiatric illness, reduce within-disorders heterogeneity, and provide clues to underlying processes and mechanisms.

Following Kraepelin (1919), development and course have been accepted as a key feature of diagnostic validity (Robins & Guze, 1970). In the fifth edition of the *Diagnostic and Statistical Manual for Mental Disorders* (American Psychiatric Association, 2013), age of onset

**Corresponding Author:**
Connor Lawhead, Department of Psychology, Stony Brook University
Email: connor.lawhead@stonybrook.edu

plays a key role in the criteria for some disorders (e.g., attention-deficit/hyperactivity disorder [ADHD], autism) and persistence is critical for others (e.g., persistent depressive disorder, schizophrenia). Yet existing classification systems fail to provide a comprehensive picture of illness progression from a life-course perspective (Maughan & Collishaw, 2015). Individuals often accumulate diagnostic comorbidities, and although disorders may persist over time, people often transition between related or unrelated disorders (Caspi et al., 2020; Copeland et al., 2013).

A life-course perspective can contribute to delineating more homogeneous groups of disorders. For example, diagnostic continuity over time provides a stronger genetic "signal" than diagnoses at a single time point (Kendler et al., 2023), the unfolding of comorbidities over time may be markers for heterogeneity within diagnostic groups (e.g., alcoholism preceded by anxiety differs in fundamental respects from alcoholism preceded by antisocial personality; Chassin et al., 2013), and the emergence and continuity of disorders at different developmental stages may reflect different conditions (e.g., adolescent-limited vs. life-course-persistent antisocial behavior; Moffitt & Caspi, 2001).

Several more recent frameworks have attempted to address these issues. The clinical-staging model (McGorry et al., 2006) takes a transdiagnostic approach, which posits that subthreshold psychopathology increases in specificity as it becomes more severe over time. However, this model does not address symptom discontinuities and the accumulation of comorbidities during the progression of illness. To address problems of heterogeneity and comorbidity, the Hierarchical Taxonomy of Psychopathology (HiTOP) is based on factor-analytic studies that decompose categorical diagnoses into more homogeneous dimensions, many of which cut across multiple disorders (Kotov et al., 2017). These symptom dimensions are aggregated into higher-order spectra, organizing psychopathology in a hierarchical fashion that provides important insights into the patterning of comorbidity. However, HiTOP does not address the emergence and course of symptom dimensions, their continuities and discontinuities, or the patterning of their interrelationships over time. Although some studies have considered the dimensionality of psychopathology across several decades (Caspi et al., 2014), the resulting factor structure did not attempt to characterize and differentiate longitudinal patterns.

Whereas factor-analytic techniques focus on relationships between variables, an alternative approach that may be more directly suited to classification focuses on relationships between people. Examples of data-driven person-centered approaches are cluster analysis, latent-profile (or latent class) analysis, and growth-mixture modeling. Cluster and latent-profile analysis have rarely been applied to longitudinal data because traditional techniques require modifications to account for their nested structure. However, unlike growth-mixture modeling, they can handle complex nonlinear relationships with a limited number of waves.

Applying data-driven person-centered approaches across a developmentally informative time frame, Healy et al. (2022) used a longitudinal extension of latent-profile analysis, latent-profile-transition analysis, to derive patterns of transitions of internalizing and externalizing symptoms in two large cohorts—one followed in childhood and another followed in adolescence. They found four profiles: no psychopathology, high levels of psychopathology, internalizing problems, and externalizing problems. About 50% of both cohorts transitioned into one of the three psychopathology profiles at some point in development, and high psychopathology was most often preceded by externalizing problems. This study robustly captured the dynamic (and in some cases, persistent) nature of psychopathology over time. However, the groupings were not truly longitudinal because they created classes cross-sectionally and examined transitions into and out of these cross-sectionally derived clusters across waves. Longitudinal clustering, in contrast, builds groups based on both cross-sectional and longitudinal variation.

Only a few studies have attempted to apply cluster or latent-profile analysis to characterize the development and/or continuity of psychopathology. Using *k*-means cluster analysis for longitudinal data, Martinek et al. (2023) clustered trajectories of weekly ratings of the course of depression for a year following hospital or clinic discharge and found five subgroups with unique patterns of recovery, relapse, and persistence in an adult sample. The PsyCourse Study (Schulte et al., 2022) focused on patients with schizophrenia and bipolar-spectrum disorders followed for 18 months. Applying the same procedure, Schulte et al. (2022) reported five distinct longitudinal clusters that differed in diagnoses and functioning based on three clinical dimensions. To our knowledge, however, no studies have applied this novel approach to a broad range of mental disorders over the course of a longer and developmentally informative time frame. By examining the trajectories of multiple disorders as they jointly coevolve over time, it is possible to identify subgroups with unique patterns of comorbidity and homotypic and heterotypic continuity during specific developmental periods.

In the present study, we applied a longitudinal clustering approach to common mental disorders in a community sample prospectively assessed triennially from early childhood through the end of adolescence. We identified subgroups of individuals based on multiple simultaneous (or joint) trajectories of diagnostic course

from ages 3 to 18 years old. We included mental disorders beginning in the preschool years, a period that has been relatively neglected in psychopathology research (Angold & Egger, 2007; Bufferd et al., 2016) but has important prognostic implications. For example, Finsaas et al. (2018) observed that 48% of preschoolers with a psychiatric diagnosis met criteria for a mental disorder in early adolescence. In using a data-driven, longitudinal clustering approach to common mental disorders throughout child and adolescent development, we accounted for the accumulation of comorbidities and continuities and transitions among disorders as an initial step toward a developmentally based classification framework. Once we identified the optimal cluster solution, we compared the cluster on a set of a priori variables typically used as predictors of later internalizing and externalizing psychopathology, including parental education and psychopathology, parenting and other markers of a child's early environment, child temperament, and child cognitive and social functioning. We also assessed how the clusters differed based on key functional outcomes, such as interpersonal and academic functioning.

## Transparency and Openness

### *Preregistration*

This study was not preregistered.

### *Data, materials, code and online resources*

All code and data are available at https://osf.io/d93ks/?view_only=158233500e6e4a0896b1ebd9fdeb4c1e.

### *Reporting*

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

### *Ethical approval*

Ethical approval was obtained by the Stony Brook University Institutional Review Board, and the study was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

## Method

### *Participants*

Participants were drawn from the Stony Brook Temperament Study,[1] a longitudinal study of risk factors and pathways to psychopathology from age 3 to age 18 (Klein & Finsaas, 2017). Families with a 3-year-old child living within 20 miles of Stony Brook, New York, were recruited using commercial mailing lists for a larger study of risk for mental disorders; children were excluded if they did not live with a biological parent or had significant medical or developmental disorders ($n = 559$). An additional 50 families were added in the second wave of assessments, when children were 6 years old, to increase the diversity of the sample. Only one child per family was included. Children were reassessed every 3 years until age 18.

Diagnostic interviews were conducted with a parent when the children were 3 ($N = 541$) and 6 ($N = 516$) years old and with a parent and the child at ages 9 ($N = 488$), 12 ($N = 476$), and 15 ($N = 458$). At age 18 ($N = 418$), only the youths were interviewed. Participants were included in the study if they had at least one wave of diagnostic information. Of these 600 participants, 45 (6.9%) completed one wave, 41 (6.3%) completed two waves, 35 (5.8%) completed three waves, 48 (8.0%) completed four waves, 113 (18.8%) completed five waves, and 318 (53.0%) completed all six assessment waves. When considering the unique participants across all samples, 272 (45.3%) were female, and 477 (87.5%) were White and non-Hispanic. The sample's demographic and socioeconomic characteristics were representative of the larger county (Bufferd et al., 2011).

### *Diagnostic assessments*

The Preschool Age Psychiatric Assessment (Egger et al., 1999), an interviewer-based, structured diagnostic interview, was administered to parents by telephone at the age-3 wave and in person at the age-6 wave. Diagnostic interviews with parents about their children conducted in person and by phone yield similar results (Lyneham & Rapee, 2005). Diagnoses, based on the fourth edition of the *Diagnostic and Statistical Manual for Mental Disorders* (*DSM-IV*; American Psychiatric Association, 1994; including modified criteria for preschool depression, Luby et al., 2002), in the past 3 months were derived following the developers' algorithms. The Preschool Age Psychiatric Assessment has good test-retest reliability over a mean 11-day interval (Egger et al., 2006). Interrater reliability in our study was assessed using audio recordings on sample of 21 interviews at age 3 and 35 interviews at age 6 enriched for psychopathology. At age 3, $\kappa$ was 1.00 for all disorders. At age 6, $\kappa$s were .64 for any depressive disorder, .89 for any anxiety disorder, .64 for ADHD, and .87 for any disruptive behavior disorder (DBD), all of which are in the moderate-substantial range (Shrout, 1998).

The Kiddie Schedule for Affective Disorders Schizophrenia Present and Lifetime Version, a semistructured

interview for school-age children and adolescents (Kaufman et al., 1997), was administered to a parent and the child at the age-9, -12, and -15 assessments and to the youth alone at the age-18 assessment. In the age-9 wave, lifetime psychopathology was ascertained, but only disorders present from ages 7 to 9 are included in our analyses. At the age-12, -15, and -18 waves, psychopathology was assessed since the previous assessment. Interrater reliabilities (indexed by $\kappa$) was determined using videotapes of interviews of samples of participants enriched for psychopathology. Interrater reliability ratings were obtained using 74 interviews at age 9, 25 interviews at ages 12 and 15, and 34 interviews at age 18. Interrater reliabilities ranged from .72 to .88 for any depressive disorder, .67 to .94 for any anxiety disorder, .85 to 1.00 for ADHD, and .58 to .91 for any DBD, all of which are in the fair, moderate, or substantial ranges (Shrout, 1998).

All diagnostic interviews were conducted in person or remotely by clinical-psychology graduate students and masters'-level clinicians supervised by a senior child and adolescent psychiatrist and clinical psychologist. In-person and remote interviews with adolescents and young adults yield comparable results (Rohde et al., 1997). The following *DSM-IV* diagnoses were examined: depressive disorders (major depressive disorder, dysthymic disorder, depressive disorder not otherwise specified [NOS] ages 3–18), anxiety disorders (specific phobia, social phobia, separation anxiety, generalized anxiety, and panic and/or agoraphobia at ages 3–18 and anxiety disorder NOS at ages 9–18), DBDs (oppositional defiant disorder and conduct disorder at ages 3–18 and DBD-NOS at ages 9–18), ADHD (ADHD at ages 3–18 and ADHD-NOS at ages 9–18), and substance use disorders (SUDs; alcohol or drug abuse or dependence at ages 9–18).

## Validation measures

The clusters in the optimal solution were validated against a set of age-3 predictors and age-18 outcomes.

### Age-3 predictors.

*Parental education and psychopathology.* Parental education, as a proxy for socioeconomic status, was defined as the number of parents with a bachelor's degree or higher. Parental psychopathology was assessed with the Structured Clinical Interview for DSM-IV nonpatient version (First & Gibbon, 2004). We determined the number of parents with a lifetime history of any depressive disorder, any anxiety disorder, and any SUD. Kappas for interrater reliability ($N$=30) were .93 for mood disorder, .91 for anxiety disorder, and 1.00 for SUD, all of which are in the substantial range (Shrout, 1998).

*Early environment.* Early environment variables consisted of the mother- and father-reported Dyadic Adjustment Scale (Spanier, 1976), a 32-item questionnaire assessing marital satisfaction (Cronbach's $\alpha$s=.94 and .95, respectively, both in the substantial range). Example items include how often parents have arguments about finances, household tasks, amount of time spent together, career decisions, and other life domains and how often they discuss divorce. In addition, the life-stress scale (Costello et al., 1998), a module of the Preschool Age Psychiatric Assessment, assesses a wide range of life events that might affect the child, including "high magnitude" events associated with posttraumatic stress disorder and "low magnitude" events (e.g., parental separation, changing schools). We summed the number of events experienced before age 3.

Regarding parenting, each child and one parent participated in a 30-min structured parent-child interaction session using a modified version of the Teaching Tasks (Egeland & Hiester, 1995). The battery consisted of six standardized tasks adapted from the Ainsworth Strange Situation Procedure (Ainsworth et al., 2015), which was designed to elicit individual differences in parenting. Tasks were video-recorded and coded for behavioral indices of support (parental expression of positive regard and emotional support), hostility (parent's expression of anger, frustration, annoyance, discounting, or rejection), and the quality of relationship between parent and child (affective and verbal sharing between child and parent, contingent responding to each other, sensitivity of parent to child's distress, and effective conflict resolution). Ratings were summed across episodes, and reliability was computed via intraclass correlations (ICCs; two-way random effects, absolute agreement) on a random sample of 55 individuals. ICCs ranged from .59 to .91, which range from fair to substantial (Shrout, 1998).

In addition, both parents completed the Parenting Styles and Dimensions Questionnaire (Robinson et al., 1995), a 47-item measure composed of three factor-analytically derived dimensions: authoritative (warmth and involvement, democratic participation), authoritarian (verbal hostility, harsh punishment), and permissive (lack of follow-through, ignoring misbehavior) parenting styles. Alphas ranged from .74 to .82, which are in the moderate-substantial range (Shrout, 1998).

*Temperament.* The Laboratory Temperament Assessment Battery (Lab-TAB; Gagne et al., 2011) is an observational measure designed to assess child temperament using a series of emotion-eliciting episodes. The child participated in 12 episodes, which were videotaped and coded for facial, vocal, and postural indicators of emotion and several emotion-relevant behaviors. Ratings were $z$-scored, summed across episodes, and used to derive

a number of temperament constructs (see Olino et al., 2010). For this article, we examined positive emotionality (positive affect and engagement/interest), negative emotionality as a whole and each of its three components (fear, sadness, anger) separately, and impulsivity. Interrater ICCs (two-way random effects, absolute agreement) for the variables included in this article, based on a random sample of 35 individuals, ranged from .73 to .89, which are in the moderate-substantial range (Shrout, 1998).

We included three episodes designed specifically to assess temperamental behavioral inhibition. The "risk room" episode had the child explore a set of novel and potentially threatening stimuli (e.g., Halloween mask, black box). The "stranger approach" episode involved a male accomplice approaching the child while left alone and speaking to the child while slowly walking closer. In the "exploring new objects" episode, the child was given the opportunity to explore ambiguous stimuli (e.g., mechanical spider). Coding procedures followed prior literature (Olino et al., 2010; Pfeifer et al., 2002). Briefly, each episode was divided into 20-s to 30-s epochs, and within each epoch, a maximum intensity rating of facial, vocal, or bodily fear was coded on a 4-point Likert scale. Behavioral inhibition was computed as the average of these standardized ratings and standardized ratings of latency to fear (reversed), latency to touch objects, total number of objects touched (reversed), tentative play, referencing the parent, proximity to parent, referencing experimenter, time spent playing (reversed), startle, sad facial affect, latency to vocalize, approach toward the stranger (reversed), avoidance of the stranger, gaze aversion, and verbal/nonverbal interaction with the stranger (reversed). Interrater ICC ($N = 28$) was .88, which is in the substantial range (Shrout, 1998).

We also assessed behavioral inhibition using parent reports. The Behavioral Inhibition Questionnaire (Bishop et al., 2003) was administered to the parent who accompanied the child to the laboratory (typically the mother). This 30-item questionnaire assesses the frequency of the child's behavioral inhibition across six contexts in the domains of social novelty, situational novelty, and novel physical activities with possible risk of injury ($\alpha = .96$). For example, items assess the child's comfort in asking other children to play, child's caution in activities that involve physical challenges, child's ability and timing in adjusting to new situations, and child's comfort with being the center of attention. We *z*-scored the Behavioral Inhibition Questionnaire and the Lab-TAB and summed them to provide a composite index of behavioral inhibition.

The child's mother completed the Children's Behavior Questionnaire (Rothbart et al., 2001), a 191-item parent-report measure designed to assess temperament in young children ($\alpha$s for 16 subscales ranged from .65 to .91, which are in the moderate-substantial range). Items comprise the following scales: activity level, anger/frustration, attentional focusing, discomfort, fear, high- and low-intensity pleasure, impulsivity, inhibitory control, perceptual sensitivity, positive anticipation, shyness, sadness, smiling/laughter, and soothability. We examined the three higher-order factors (surgency, negative affectivity, effortful control) derived from Rothbart et al.'s (2001) factor analyses of the Children's Behavior Questionnaire subscales.

*Cognitive and social functioning.* To index cognitive functioning, each child completed the third edition of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997) and the Expressive One-Word Vocabulary Test (Brownell, 2000); these tests assess receptive and expressive vocabulary, respectively, based on presentation of various photos (e.g., children are asked to pick the square out of a series of photos of shapes). In addition, to index social functioning, the 15-item Social Competence Scale from the Vineland Screener (Sparrow et al., 1993), a parent-report measure of children's adaptive behaviors, such as communication, socialization, and daily living skills, was administered. Example items include the child's ability to make eye contact when meeting new people, child's caution around things that could burn him or her, and child's ability to pay attention to 15-min stories.

***Age-18 functioning outcomes.*** The UCLA Life Events Interview (Hammen et al., 1987) was administered to participants in the age-18 wave. Although the interview was designed to assess episodic and chronic stress, the latter scores can readily be interpreted as reflecting functional impairment (Harkness & Monroe, 2016). Interviewers used behavioral probes to assess functioning over the past year on a 5-point scale, including half points (higher scores indicate poorer functioning). For the present study, we examined academic/work functioning and interpersonal (family, friends, peers, and romantic partners) functioning. Interrater ICCs for each domain ($N = 34$) included in these summary scores ranged from .65 to .89, which are in the moderate-substantial range (Shrout, 1998).

## *Data analysis*

***Cluster estimation.*** The *k*-means cluster modeling for longitudinal data (*K-means longitudinal 3D* [*kml3d*]; Genolini et al., 2013) package using R (Version 4.2.0) was used to identify distinct clusters of psychopathology trajectories over six assessment time points. The *kml3d* package offers a nonparametric, expectation-maximization algorithm that clusters joint variable-trajectories, capturing the evolution and complex interactions between

variables over time (Genolini et al., 2015). To evaluate the optimal number of cluster trajectories, we tested six longitudinal *k*-means models, increasing the number of clusters stepwise from two to seven (computed using the *kml3d* algorithm). The *k*-means algorithm was initialized with the following procedure:

> a) choose one center $c_0$ uniformly at random from among the data points; b) for each data point $x$, compute $D(x)$, the distance between $x$ and $c_0$; c) choose one new center $c_1$ at random using a weighted probability distribution proportional to $D(x)^2$; d) remove $c_0$ from the list of centers; e) for each data point $x$, compute $D(x)$, the distance between x and the nearest center that has already been chosen; f) randomly choose a data point as the new center $c_i$, using a weighted probability distribution where a point $x$ is chosen with probability proportional to $D(x)^2$; g) repeat steps e and f until k centers have been chosen. (Genolini et al., 2015)

The algorithm was run 50 times, a maximum of 500 iterations were run if convergence was not reached, and individual runs were automatically sorted by best fit. As part of the model, we imputed missing data using linear interpolation and then added a variation to make the trajectory follow the "shape" of the population's mean trajectory; thus, overall trends are informing cluster proportions and within-persons changes across time (Genolini et al., 2013). Because diagnostic data are binary, we used the deviance distance metric (rather than the Euclidean distance metric for continuous data) to find the optimal number of centroids. We assessed model fit using the Caliński-Harabasz index (Caliński & Harabasz, 1974), Genolini variant. This variant is notated as $C_G(k) = (Trace(B) / Trace(W)) \times (n - k / \sqrt{k} - 1)$, where $B$ is the between-clusters covariance matrix and $W$ is the within-clusters covariance matrix. High values of $Trace(B)$ denote well-separated clusters, and low values of $Trace(W)$ denote compact clusters (Genolini et al., 2015). This variant has the advantage of jointly considering both parsimony and fit. The Caliński-Harabasz index was found to be the best index in detecting the optimal number of clusters (Milligan & Cooper, 1985).

***Cluster validation.*** Once clusters were estimated and extracted from *kml3d*, handling of missing values in the validation measures was performed using multiple imputation using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011). Twenty imputations were conducted; 100 was the maximum number of iterations if convergence was not reached. The reverse monotone visit sequence was used so that data from the variables with the greatest amount of missingness were imputed first. Imputation performance was assessed by inspecting density plots of all imputations at once to ensure they followed the same shape of distribution.

**Table 1.** Fit Indices for Best-Fitting Solution for Each *k*-Means Model

| Cluster solution | Best-fitting Caliński-Harabsz (Genolini variant) fit index |
|---|---|
| 2 | 105.104 |
| 3 | 144.631 |
| 4 | 173.115 |
| 5 | 193.703 |
| 6 | 198.020 |
| 7 | 193.796 |

For descriptive purposes, chi-square tests were performed at each imputation to assess differences in the prevalence of the diagnostic categories among the clusters at each time point; these values were then pooled across imputations. Chi-square tests were also performed on demographic variables (i.e., sex and race/ethnicity). Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander) and ethnicity (Hispanic or non-Hispanic) were measured via self-report. For validation purposes, one-way analyses of variance (ANOVAs) and post hoc Holm-Bonferroni-corrected pairwise *t* tests between the "healthy" (reference) cluster and every other cluster were computed on each imputed data set, which were then pooled. Corresponding Cohen's *d* values were computed for each comparison.

## Results

Table 1 displays respective Caliński-Harabasz (Genolini variant) fit indices for the best-fitting cluster solutions across iterations for the two- to seven-cluster solutions; the six-cluster solution provides the best fit. Figure 1 displays the trajectories of the six clusters as a function of each of the five disorders. Cluster A (46.7% of the sample, $n = 280$), demarcated by the red line in Figure 1, is characterized primarily by the absence of psychopathology from early childhood through adolescence (i.e., the healthy cluster). Cluster B (13.5% of the sample, $n = 81$), demarcated by the yellow line in Figure 1, is characterized by elevated rates of anxiety disorders, particularly in late childhood and early adolescence, and is referred to as the "anxiety" (ANX) cluster. Cluster C (12.3%, $n = 74$), demarcated by the green line in Figure 1, is characterized by increasing rates of ADHD through early adolescence and is referred to as the "ADHD" cluster. Cluster D (12%, $n = 72$), demarcated by

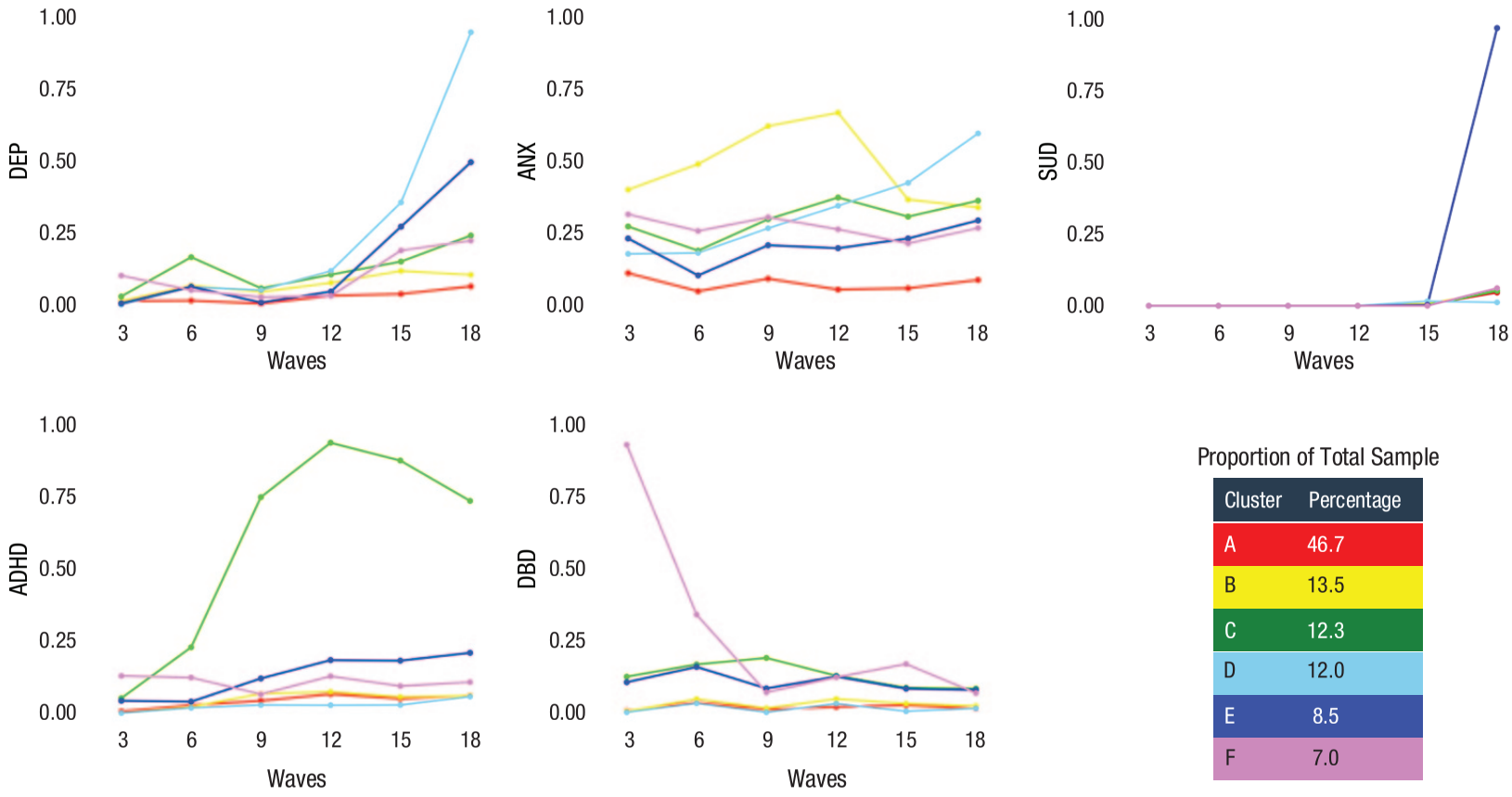


**Fig. 1.** Cluster visualization for best-fitting cluster solution. DEP = any depressive disorder, including not otherwise specified; ANX = any anxiety disorder, including not otherwise specified; ADHD = attention-deficit/hyperactivity disorder; DBD = disruptive behavior disorder; SUD = substance use disorder.

the light blue line in Figure 1, is characterized by gradually rising rates of anxiety disorders throughout childhood and adolescence and a rapid increase in rates of depressive disorders beginning in mid-adolescence. We refer to this cluster as the "depression/anxiety" (DEP/ANX) cluster. Cluster E (8.5%, $n = 51$), demarcated by the dark blue line in Figure 1, is characterized by moderately rising rates of depressive disorders through adolescence and a very sharp increase in SUD in late adolescence. This cluster is referred to as the "SUD/depression" (SUD/DEP) cluster. Finally, Cluster F (7%, $n = 42$), demarcated by the pink line in Figure 1, is characterized by high but decreasing rates of DBDs through childhood and a modest increase in mid-adolescence. This cluster is referred to as the "DBD" cluster.

Qualitatively, this optimal six-cluster solution differed from the five-cluster solution in that the five-cluster solution did not capture childhood anxiety as its own cluster. Instead, most participants in the Cluster B (ANX) group were included in Cluster A (the healthy cluster). The seven-cluster solution, by contrast, included two clusters characterized by DBD—one cluster with moderate rates of DBD throughout childhood and adolescence often accompanied by ADHD ($n = 22$) and the other cluster with a spike in DBD in adolescence, rising rates of ADHD in adolescence, and moderate and increasing rates of anxiety across childhood and adolescence ($n = 17$). Thus, although the seven-cluster solution is theoretically interesting, the overall model fit is poorer than that of the six-cluster solution, and the $n$ for some clusters was quite small.

Table 2 shows the comparison of the distribution of diagnoses and chi-square results at each wave across the six clusters. Except for SUD at age 15 (because only one individual had this diagnosis), all omnibus chi-square values for all disorder categories at all time points were significant. Post hoc Holm-Bonferroni-corrected multiple comparisons show that Cluster B (ANX) had significantly higher rates of anxiety disorders than all other clusters at ages 9 and 12. Cluster C (ADHD) had significantly higher rates of ADHD than all other clusters at ages 9, 12, 15, and 18. Cluster D (DEP/ANX) had significantly higher rates of depression than all other clusters at age 18 and significantly higher rates of anxiety than all other clusters (except Cluster C) at age 18. Cluster E (DEP/SUD) also had significantly higher rates of depression than all other clusters (except Cluster D) at age 18 and significantly higher rates of SUD than all other clusters at age 18. Cluster F (DBD) had significantly higher rates of DBDs than all other clusters at age 3.

We also examined how clusters differed on sex and race/ethnicity (see Table 3). Proportion of males to females significantly differed across clusters. Holm-Bonferroni-corrected post hoc multiple comparisons revealed that Cluster C (ADHD) had significantly more males than females and that Cluster D (DEP/ANX) had significantly more females than males. The clusters did not differ on race or ethnicity; however, these results are underpowered given the relatively homogeneous White, non-Hispanic study sample.

Next, we examined the associations with the external validators to assess validity of the clusters. The external validators included a series of predictors from the initial (age 3) wave and functional outcomes at the final (age 18) wave. Table 3 displays the means and standard deviations of the validation measures for each cluster and the results from one-way ANOVAs. All ANOVA results were significant except positive emotionality from the Lab-TAB. We continued to examine pairwise $t$ tests for all variables because it is possible to detect meaningful group differences despite a nonsignificant omnibus test (Tian et al., 2018). Table 4 displays corresponding Holm-Bonferroni-corrected $t$ tests and corresponding $p$ values and effect sizes (using Cohen's $d$) comparing Cluster A (healthy) with each of the other clusters. Significant effect sizes were in the medium-to-large range. For pairwise cluster comparisons with each psychopathology cluster as the reference group, see the Supplemental Material available online. A heatmap of effect sizes are displayed in Figure 2.

## *Age-3 predictors*

Comparing Cluster B (ANX) with Cluster A (healthy), we found that Cluster B had significantly higher levels of father-reported dyadic adjustment and higher levels of observed negative emotionality, anger, sadness, and behavioral inhibition. Cluster B also had significantly higher levels of mother-reported negative affect and significantly lower levels of surgency.

By contrast, compared with cluster A (healthy), Cluster C (ADHD) had parents with significantly higher rates of anxiety disorders, lower mother-reported dyadic adjustment, and higher parent-reported life stress affecting the child. In addition, this cluster had a poorer observed quality of relationship with the parent and higher father-reported authoritarian parenting. Cluster C also had lower receptive-vocabulary scores and lower levels of parent-reported social competence on the Vineland Screener. This cluster exhibited lower levels of observed positive emotionality, higher observed impulsivity, higher mother-reported surgency, and lower mother-reported effortful control.

Cluster D (DEP/ANX) had parents with significantly higher rates of mood disorders and higher expressive-vocabulary scores on the Expressive One-Word Vocabulary Test.

**Table 2.** Chi-Square Comparison of Diagnoses Among Best-Fitting Cluster Solution

| | | A (healthy) $n=280$ | | B (ANX) $n=81$ | | C (ADHD) $n=74$ | | D (DEP/ANX) $n=72$ | | E (SUD/DEP) $n=51$ | | F (DBD) $n=42$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diagnosis | Age | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | $\chi^2(5)$ |
| DEP | 3 | $4_{cde}$ | 1.34 | $1_{cdf}$ | 1.35 | $2_{abdef}$ | 2.96 | $0_{abcef}$ | 0.07 | $0_{cdf}$ | 0.56 | $4_{abcde}$ | 10.20 | 18.77** |
| | 6 | $4_{bcdef}$ | 1.51 | $5_{ac}$ | 6.76 | $12_{abdef}$ | 16.60 | $4_{ac}$ | 6.22 | $3_{ac}$ | 6.35 | $2_{ac}$ | 5.10 | 27.78*** |
| | 9 | $1_{bcdf}$ | 0.49 | $4_{ae}$ | 4.53 | $4_{ae}$ | 5.79 | $4_{ae}$ | 5.16 | $0_{bcdf}$ | 0.84 | $1_{ae}$ | 2.72 | 14.22* |
| | 12 | $9_{bcd}$ | 3.21 | $6_{adef}$ | 7.70 | $8_{aef}$ | 10.55 | $9_{abef}$ | 11.84 | $2_{bcd}$ | 4.67 | $1_{bcd}$ | 3.17 | 13.50* |
| | 15 | $11_{bcdef}$ | 3.81 | $10_{acdef}$ | 11.82 | $11_{abdef}$ | 15.06 | $26_{abcef}$ | 35.58 | $14_{abcd}$ | 27.17 | $8_{abcd}$ | 18.93 | 64.14*** |
| | 18 | $18_{bcdef}$ | 6.46 | $8_{acdef}$ | 10.46 | $18_{abdef}$ | 24.13 | $68_{abcef}$ | 94.71 | $25_{abcdf}$ | 49.58 | $9_{abcde}$ | 22.34 | 265.05*** |
| ANX | 3 | $31_{bcdef}$ | 11.24 | $33_{acdef}$ | 40.15 | $20_{abdf}$ | 27.41 | $13_{abcef}$ | 17.92 | $12_{abdf}$ | 23.25 | $13_{abcde}$ | 31.63 | 40.56*** |
| | 6 | $14_{bcdef}$ | 4.97 | $40_{acdef}$ | 49.03 | $14_{abef}$ | 19.05 | $13_{abef}$ | 18.25 | $5_{abcdf}$ | 10.46 | $11_{abcde}$ | 25.85 | 96.51*** |
| | 9 | $26_{bcdef}$ | 9.32 | $50_{acdef}$ | 62.20 | $22_{abe}$ | 29.92 | $19_{abe}$ | 26.79 | $11_{abcdf}$ | 20.92 | $13_{abe}$ | 30.61 | 103.81*** |
| | 12 | $16_{bcdef}$ | 5.54 | $54_{acdef}$ | 66.84 | $28_{abef}$ | 37.45 | $25_{abcf}$ | 34.59 | $10_{abcdf}$ | 19.89 | $11_{abcde}$ | 26.42 | 147.97*** |
| | 15 | $17_{bcdef}$ | 6.00 | $30_{acdf}$ | 36.74 | $23_{abdef}$ | 30.82 | $31_{abcef}$ | 42.53 | $12_{acd}$ | 23.25 | $9_{abcd}$ | 21.54 | 77.63*** |
| | 18 | $25_{bcdef}$ | 8.86 | $28_{acdf}$ | 34.04 | $27_{abd}$ | 36.36 | $43_{abcef}$ | 59.66 | $15_{ad}$ | 29.51 | $11_{abd}$ | 26.87 | 95.94*** |
| ADHD | 3 | $2_{cef}$ | 0.63 | $0_{cef}$ | 0.06 | $4_{abdf}$ | 5.21 | $0_{cef}$ | 0.07 | $2_{abdf}$ | 4.30 | $5_{abcde}$ | 12.93 | 33.76*** |
| | 6 | $8_{cef}$ | 2.76 | $2_{cef}$ | 2.18 | $17_{abdef}$ | 22.84 | $1_{cef}$ | 1.85 | $2_{abcdf}$ | 4.01 | $5_{abcde}$ | 12.36 | 52.10*** |
| | 9 | $12_{bcef}$ | 4.30 | $5_{acde}$ | 6.76 | $55_{abdef}$ | 74.84 | $2_{bcef}$ | 2.85 | $6_{abcdf}$ | 12.04 | $3_{acde}$ | 6.58 | 260.99*** |
| | 12 | $19_{bcdef}$ | 6.62 | $6_{acdef}$ | 7.41 | $69_{abdef}$ | 93.69 | $2_{abcef}$ | 2.78 | $9_{abcdf}$ | 18.39 | $5_{abcde}$ | 12.81 | 322.88*** |
| | 15 | $14_{bcef}$ | 5.02 | $5_{acdef}$ | 5.64 | $65_{abdef}$ | 87.52 | $2_{bcef}$ | 2.84 | $9_{abcdf}$ | 18.21 | $4_{abcde}$ | 9.41 | 317.07*** |
| | 18 | $17_{cef}$ | 6.12 | $5_{cef}$ | 6.00 | $54_{abdef}$ | 73.55 | $4_{cef}$ | 5.75 | $11_{abcdf}$ | 20.92 | $4_{abcde}$ | 10.77 | 221.49*** |
| DBD | 3 | $2_{bcef}$ | 0.65 | $0_{acef}$ | 0.47 | $9_{abdf}$ | 12.48 | $0_{cdef}$ | 0.13 | $5_{abdf}$ | 10.55 | $39_{abcde}$ | 92.97 | 388.24*** |
| | 6 | $10_{bcef}$ | 3.52 | $4_{acdef}$ | 4.70 | $12_{abdf}$ | 16.73 | $2_{abcef}$ | 3.24 | $8_{abdf}$ | 15.87 | $14_{abcde}$ | 34.01 | 58.93*** |
| | 9 | $3.10_{bcef}$ | 1.11 | $1_{acdef}$ | 1.53 | $14_{abdef}$ | 18.98 | $0.10_{bcef}$ | 0.13 | $4.29_{abcd}$ | 8.40 | $2_{abcd}$ | 7.03 | 54.13*** |
| | 12 | $6_{bcdef}$ | 1.99 | $4_{acef}$ | 4.76 | $9.48_{abd}$ | 12.81 | $2_{acef}$ | 3.11 | $6_{abd}$ | 12.51 | $5_{abd}$ | 12.24 | 24.75*** |
| | 15 | $7_{cdef}$ | 2.60 | $3_{cdef}$ | 3.12 | $6.38_{abdf}$ | 8.62 | $0_{abcef}$ | 0.46 | $4_{abdf}$ | 8.31 | $7_{abcde}$ | 16.87 | 23.80*** |
| | 18 | $4_{bcdef}$ | 1.34 | $2_{acef}$ | 2.29 | $6.19_{abd}$ | 8.37 | $1_{acef}$ | 1.46 | $4_{abd}$ | 7.84 | $3_{abd}$ | 6.80 | 17.61** |
| SUD | 15 | 1 | 0.48 | 1 | 0.65 | 0.19 | 0.26 | 1 | 1.59 | 0 | 0.28 | 0 | 0.00 | 4.60 |
| | 18 | $13_{cdef}$ | 4.57 | $4_{cef}$ | 5.23 | $3.86_{abef}$ | 5.21 | $1_{aef}$ | 1.19 | $49_{abcdf}$ | 96.64 | $3_{abcde}$ | 6.12 | 378.75*** |

Note: Estimates reflect pooled raw numbers of number of participants with the disorder in each cluster. Percentages reflect proportion of individuals with the disorder within each cluster. Squared values indicate significant difference between proportions (Holm-Bonferroni-corrected for multiple comparisons). Means with different subscripts differ at the $p=.05$ level. DEP=any depressive disorder, including not otherwise specified; ANX=any anxiety disorder, including not otherwise specified; ADHD = attention-deficit/hyperactivity disorder; DBD=disruptive behavior disorder; SUD=substance use disorder.
*$p<.05$. **$p<.01$. ***$p<.001$.

**Table 3.** Means, Standard Deviations (or Percentages and Sample Sizes for Categorical Variables), and One-Way Analysis of Variance of Each Measure by Each Cluster

| Measure | A (healthy) | B (ANX) | C (ADHD) | D (DEP/ANX) | E (SUD/DEP) | F (DBD) | $\chi^2(5, N=600)$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| Percentage female | 44.3 (124) | 51.9 (42) | 28.4 (21) | 73.6 (53) | 37.3 (19) | 40.0 (13) | 38.18 | < .001 |
| Percentage non-White | 12.1 (34) | 8.6 (7) | 16.2 (12) | 2.8 (2) | 15.7 (8) | 9.5 (4) | 23.30 | .274 |
| Percentage Hispanic | 13.2 (37) | 8.6 (7) | 10.8 (8) | 9.7 (7) | 21.6 (11) | 11.9 (5) | 5.78 | .328 |
| Age-3 predictors | | | | | | | $F(5, 594)$ | $\eta^2$ |
| Percentage of parents with college degree | 52.5 (147) | 51.0 (41) | 43.3 (32) | 54.0 (39) | 36.3 (18) | 42.9 (18) | 6.99*** | .14 |
| Percentage of parents with lifetime mood disorder | 19.7 (55) | 29.6 (24) | 28.8 (21) | 32.4 (23) | 23.7 (12) | 34.3 (14) | 6.24*** | .09 |
| Percentage of parents with lifetime anxiety disorder | 22.3 (62) | 29.8 (24) | 35.5 (26) | 29.8 (21) | 25.4 (13) | 32.9 (14) | 5.14*** | .09 |
| Percentage of parents with lifetime substance use disorder | 26.4 (74) | 29.2 (24) | 35.8 (26) | 31.5 (23) | 41.0 (21) | 29.3 (12) | 4.68*** | .09 |
| Mother-reported dyadic adjustment | 16.34 (3.76) | 16.65 (3.21) | 14.77 (4.21) | 15.81 (3.63) | 15.43 (3.88) | 14.86 (4.22) | 3.51** | .03 |
| Father-reported dyadic adjustment | 16.32 (3.52) | 16.81 (3.09) | 15.19 (3.87) | 15.79 (3.51) | 16.51 (3.32) | 15.29 (3.32) | 2.61* | .02 |
| Parent-report life stress of child | 2.39 (1.63) | 2.79 (1.68) | 3.69 (1.85) | 2.76 (1.76) | 3.31 (1.90) | 2.98 (2.08) | 3.67** | .01 |
| Teaching Task support | 4.48 (0.55) | 4.40 (0.75) | 4.34 (0.65) | 4.54 (0.45) | 4.47 (0.54) | 4.32 (0.78) | 4.44** | .10 |
| Teaching Task hostility | 1.19 (0.30) | 1.22 (0.44) | 1.23 (0.33) | 1.12 (0.21) | 1.19 (0.36) | 1.37 (0.52) | 9.92*** | .26 |
| Teaching Task quality of relationship | 4.00 (0.53) | 3.98 (0.71) | 3.80 (0.65) | 4.08 (0.52) | 4.02 (0.58) | 3.78 (0.86) | 6.34*** | .10 |
| PSDQ mother-reported authoritarian parenting | 19.50 (3.97) | 19.45 (4.01) | 20.90 (4.21) | 18.81 (3.95) | 20.28 (4.91) | 23.66 (4.60) | 9.60*** | .07 |
| PSDQ father-reported authoritarian parenting | 19.85 (4.43) | 20.25 (4.72) | 21.88 (4.83) | 19.13 (4.10) | 19.96 (4.82) | 23.56 (4.54) | 7.73*** | .06 |
| PSDQ mother-reported permissive parenting | 10.48 (2.97) | 10.51 (3.21) | 11.27 (3.18) | 9.83 (2.75) | 11.22 (3.57) | 13.61 (3.55) | 9.56*** | .07 |
| PSDQ father-reported permissive parenting | 10.72 (2.94) | 11.42 (2.95) | 11.38 (3.13) | 11.22 (3.36) | 12.05 (3.27) | 12.71 (4.00) | 4.16** | .03 |

*(continued)*

**Table 3.** *(continued)*

| Measure | A (healthy) | B (ANX) | C (ADHD) | D (DEP/ANX) | E (SUD/DEP) | F (DBD) | $F(5, 594)$ | $\eta^2$ |
|---|---|---|---|---|---|---|---|---|
| Peabody Picture Vocabulary Test | 102.60 (14.53) | 104.48 (14.97) | 98.21 (13.96) | 105.16 (12.38) | 102.38 (12.27) | 103.83 (11.68) | 2.32* | .02 |
| Expressive One-Word Vocabulary Test | 100.73 (12.64) | 100.10 (12.66) | 97.80 (15.14) | 105.01 (11.90) | 101.09 (11.99) | 95.46 (13.33) | 3.75** | .03 |
| Vineland Screener | 19.40 (3.65) | 18.59 (3.77) | 17.08 (3.49) | 20.19 (3.55) | 19.18 (3.39) | 16.75 (3.49) | 9.87*** | .08 |
| Lab-TAB positive emotionality | 0.171 (1.86) | –0.142 (2.07) | –0.473 (1.63) | –0.161 (1.66) | –0.168 (1.72) | –0.162 (1.59) | 1.73 | .01 |
| Lab-TAB negative emotionality | 0.547 (0.260) | 0.631 (0.333) | 0.584 (0.246) | 0.561 (0.286) | 0.623 (0.220) | 0.620 (0.283) | 10.70*** | .36 |
| Lab-TAB anger | 0.555 (0.333) | 0.652 (0.388) | 0.645 (0.350) | 0.505 (0.338) | 0.564 (0.310) | 0.710 (0.365) | 11.83*** | .26 |
| Lab-TAB sadness | 0.532 (0.312) | 0.672 (0.409) | 0.565 (0.274) | 0.569 (0.317) | 0.540 (0.292) | 0.587 (0.304) | 10.48*** | .29 |
| Lab-TAB fear | 0.658 (0.345) | 0.658 (0.352) | 0.640 (0.363) | 0.702 (0.389) | 0.825 (0.344) | 0.657 (0.363) | 8.53*** | .25 |
| Lab-TAB global impulsivity | 0.675 (0.320) | 0.604 (0.335) | 0.813 (0.350) | 0.593 (0.290) | 0.713 (0.332) | 0.904 (0.366) | 18.56*** | .27 |
| Behavioral inhibition $z$ score | –0.210 (1.54) | 0.804 (1.66) | –0.292 (1.72) | 0.154 (1.49) | 0.105 (1.50) | 0.199 (1.61) | 3.50** | .02 |
| CBQ surgency | 4.81 (0.65) | 4.44 (0.84) | 5.08 (0.71) | 4.79 (0.67) | 4.98 (0.62) | 5.13 (0.72) | 9.54*** | .08 |
| CBQ negative affectivity | 3.82 (0.55) | 4.08 (0.54) | 3.98 (0.62) | 3.95 (0.60) | 3.84 (0.58) | 4.25 (0.48) | 9.92*** | .11 |
| CBQ effortful control | 5.09 (0.49) | 4.99 (0.52) | 4.64 (0.62) | 5.23 (0.57) | 4.87 (0.63) | 4.70 (0.60) | 14.31*** | .11 |
| Age-18 functional outcomes | | | | | | | | |
| LSI Chronic Academic Stress | 1.81 (0.35) | 1.82 (0.41) | 2.07 (0.48) | 1.91 (0.44) | 2.30 (0.63) | 1.78 (0.36) | 18.32*** | .18 |
| LSI Chronic Interpersonal Stress (parent included) | 2.03 (0.34) | 2.06 (0.31) | 2.21 (0.44) | 2.21 (0.46) | 2.39 (0.46) | 2.22 (0.48) | 17.13*** | .21 |

Note: Standard deviations or sample sizes for categorical variables are listed in parentheses. PSDQ=Parenting Styles and Dimensions Questionnaire; Lab-TAB=Laboratory Temperament Assessment Battery; CBQ=Children's Behavior Questionnaire; LSI=Life Stress Interview.
*$p<.05$. **$p<.01$. ***$p<.001$.

**Table 4.** Post Hoc Comparisons of All Psychopathology Clusters With Cluster A

| Measure | B (ANX) | | | C (ADHD) | | | D (DEP/ANX) | | | E (DEP/SUD) | | | F (DBD) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age 3 predictors | *t*(359) | *p* | Cohen's *d* | *t*(352) | *p* | Cohen's *d* | *t*(350) | *p* | Cohen's *d* | *t*(329) | *p* | Cohen's *d* | *t*(320) | *p* | Cohen's *d* |
| Number of parents with bachelor's degree | −0.398 | .693 | 0.049 | −1.86 | .071 | 0.240 | 0.116 | .818 | 0.031 | **−3.06** | **.003** | **0.463** | −1.59 | .121 | 0.259 |
| Number of parents with lifetime mood disorder | 1.92 | .056 | 0.261 | 1.86 | .063 | 0.256 | **2.67** | **.008** | **0.369** | 0.296 | .768 | 0.050 | 2.26 | .024 | 0.399 |
| Number of parents with lifetime anxiety disorder | 2.15 | .032 | 0.280 | **3.55** | **< .001** | **0.461** | 1.53 | .128 | 0.208 | 0.983 | .326 | 0.158 | 1.54 | .123 | 0.267 |
| Number of parents with lifetime substance use disorder | 0.401 | .689 | 0.053 | 1.94 | .052 | 0.259 | 0.890 | .374 | 0.122 | **3.11** | **.002** | **0.486** | 0.226 | .821 | 0.039 |
| Mother-reported dyadic adjustment | 1.53 | .126 | 0.203 | **−3.45** | **< .001** | **0.447** | −0.664 | .507 | 0.090 | −1.44 | .150 | 0.222 | −2.40 | .017 | 0.397 |
| Father-reported dyadic adjustment | **2.94** | **.003** | **0.369** | −2.45 | .014 | 0.305 | −0.051 | .959 | 0.007 | 1.68 | .093 | 0.251 | −2.21 | .028 | 0.358 |
| Parent-report life stress of child | 1.25 | .213 | 0.166 | **5.02** | **< .001** | **0.678** | 1.40 | .162 | 0.194 | **3.07** | **.002** | **0.484** | 2.03 | .044 | 0.342 |
| Teaching Tasks support | −1.08 | .279 | 0.132 | −1.45 | .147 | 0.193 | 0.720 | .472 | 0.104 | 0.250 | .803 | 0.040 | −1.56 | .120 | 0.256 |
| Teaching Tasks hostility | 0.562 | .574 | 0.069 | −0.75 | .455 | 0.103 | −1.52 | .128 | 0.230 | −0.245 | .807 | 0.039 | **2.96** | **.003** | **0.473** |
| Teaching Tasks quality of relationship | −0.888 | .375 | 0.118 | **−2.71** | **.007** | **0.385** | 0.739 | .460 | 0.112 | 0.019 | .985 | 0.003 | **−2.79** | **.005** | **0.478** |

*(continued)*

**Table 4.** *(continued)*

| | B (ANX) | | | C (ADHD) | | | D (DEP/ANX) | | | E (DEP/SUD) | | | F (DBD) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Measure | *t*(359) | *p* | Cohen's *d* | *t*(352) | *p* | Cohen's *d* | *t*(350) | *p* | Cohen's *d* | *t*(329) | *p* | Cohen's *d* | *t*(320) | *p* | Cohen's *d* |
| PSDQ mother-reported authoritarian parenting | 0.092 | .927 | 0.011 | 1.93 | .054 | 0.254 | −1.39 | .164 | 0.188 | 1.57 | .117 | 0.234 | **5.90** | **< .001** | **0.973** |
| PSDQ father-reported authoritarian parenting | 1.27 | .217 | 0.157 | **3.23** | **.001** | **0.424** | −1.31 | .190 | 0.180 | 4.45 | .656 | 0.068 | **4.35** | **< .001** | **0.734** |
| PSDQ mother-reported permissive parenting | −0.043 | .966 | 0.006 | 2.02 | .044 | 0.271 | −1.65 | .099 | 0.231 | 2.03 | .043 | 0.313 | **6.38** | **< .001** | **1.07** |
| PSDQ father-reported permissive parenting | 1.93 | .054 | 0.261 | 1.63 | .104 | 0.226 | 0.704 | .482 | 0.097 | **3.50** | **<.001** | **0.563** | **3.33** | **< .001** | **0.562** |
| Peabody Picture Vocabulary Test | 1.20 | .229 | 0.143 | **−2.73** | **.007** | **0.341** | 1.53 | .128 | 0.197 | −0.024 | .981 | 0.004 | 1.09 | .278 | 0.174 |
| Expressive One-Word Vocabulary Test | −0.163 | .871 | 0.021 | −1.98 | .048 | 0.249 | **2.69** | **.008** | **0.360** | 0.363 | .717 | 0.056 | **−2.50** | **.012** | **0.412** |
| Vineland Screener | −1.66 | .098 | 0.205 | **−4.70** | **< .001** | **0.611** | 1.74 | .083 | 0.228 | −0.630 | .529 | 0.096 | **−4.67** | **< .001** | **0.767** |
| Lab-TAB positive emotionality | −1.42 | .155 | 0.173 | **−3.29** | **.001** | **0.435** | −1.53 | .126 | 0.204 | −2.07 | .039 | 0.314 | −0.459 | .647 | 0.076 |
| Lab-TAB negative emotionality | **2.85** | **.005** | **0.356** | 1.48 | .140 | 0.207 | 0.700 | .485 | 0.096 | 2.28 | .023 | 0.375 | 1.47 | .141 | 0.256 |
| Lab-TAB anger | **2.48** | **.012** | **0.310** | 2.29 | .023 | 0.304 | −0.952 | .342 | 0.129 | −0.396 | .692 | 0.063 | **2.61** | **.009** | **0.440** |
| Lab-TAB sadness | **3.73** | **< .001** | **0.462** | 1.22 | .223 | 0.172 | 0.768 | .443 | 0.107 | 1.11 | .268 | 0.180 | 0.819 | .413 | 0.144 |
| Lab-TAB fear | 0.345 | .730 | 0.045 | −0.300 | .765 | 0.040 | 1.38 | .167 | 0.185 | **3.50** | **< .001** | **0.553** | 0.191 | .849 | 0.033 |

*(continued)*

**Table 4.** *(continued)*

| | B (ANX) | | | C (ADHD) | | | D (DEP/ANX) | | | E (DEP/SUD) | | | F (DBD) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Measure | *t*(359) | *p* | Cohen's *d* | *t*(352) | *p* | Cohen's *d* | *t*(350) | *p* | Cohen's *d* | *t*(329) | *p* | Cohen's *d* | *t*(320) | *p* | Cohen's *d* |
| Lab-TAB global impulsivity | −1.80 | .073 | 0.230 | **3.50** | **< .001** | **0.461** | −1.85 | .065 | 0.256 | −0.121 | .903 | 0.019 | **4.18** | **< .001** | **0.700** |
| Behavioral inhibition *z* score | **5.52** | **< .001** | **0.705** | 0.041 | .968 | 0.005 | 2.06 | .040 | 0.281 | 1.78 | .076 | 0.280 | 1.84 | .067 | 0.311 |
| CBQ surgency | **−3.93** | **< .001** | **0.501** | **2.99** | **.003** | **0.414** | 0.437 | .744 | 0.046 | 1.05 | .296 | 0.174 | **2.88** | **.004** | **0.508** |
| CBQ negative affectivity | **3.12** | **.002** | **0.422** | 2.06 | .039 | 0.281 | 1.41 | .159 | 0.196 | −0.084 | .933 | 0.014 | **4.71** | **< .001** | **0.845** |
| CBQ effortful control | −1.98 | .049 | 0.270 | **−6.43** | **< .001** | **0.877** | 1.73 | .083 | 0.245 | **−3.76** | **< .001** | **0.603** | **−4.79** | **< .001** | **0.848** |
| Age-18 functional outcomes | | | | | | | | | | | | | | | |
| LSI Chronic Academic/Work Stress | 1.31 | .191 | 0.180 | **4.99** | **< .001** | **0.683** | 2.08 | .038 | 0.296 | **8.09** | **< .001** | **1.21** | 0.269 | .788 | 0.050 |
| LSI Chronic Interpersonal Stress | 1.20 | .231 | 0.172 | **4.05** | **< .001** | **0.553** | **4.56** | **< .001** | **0.624** | **6.48** | **< .001** | **1.04** | **3.31** | **< .001** | **0.578** |

Note: Independent samples *t* test comparing parameter estimates between Cluster A and each psychopathology cluster. Holm-Bonferroni-corrected significant *p* values are bold. Red cells indicate significantly higher than Cluster A, and blue cells indicate significantly lower than Cluster A. Cohen's *d* ranges are small = 0.2–0.5, medium = 0.5–0.8, and large = 0.8–1.20. PSDQ=Parenting Styles and Dimensions Questionnaire; Lab-TAB=Laboratory Temperament Assessment Battery; CBQ=Children's Behavior Questionnaire; LSI=Life Stress Interview.

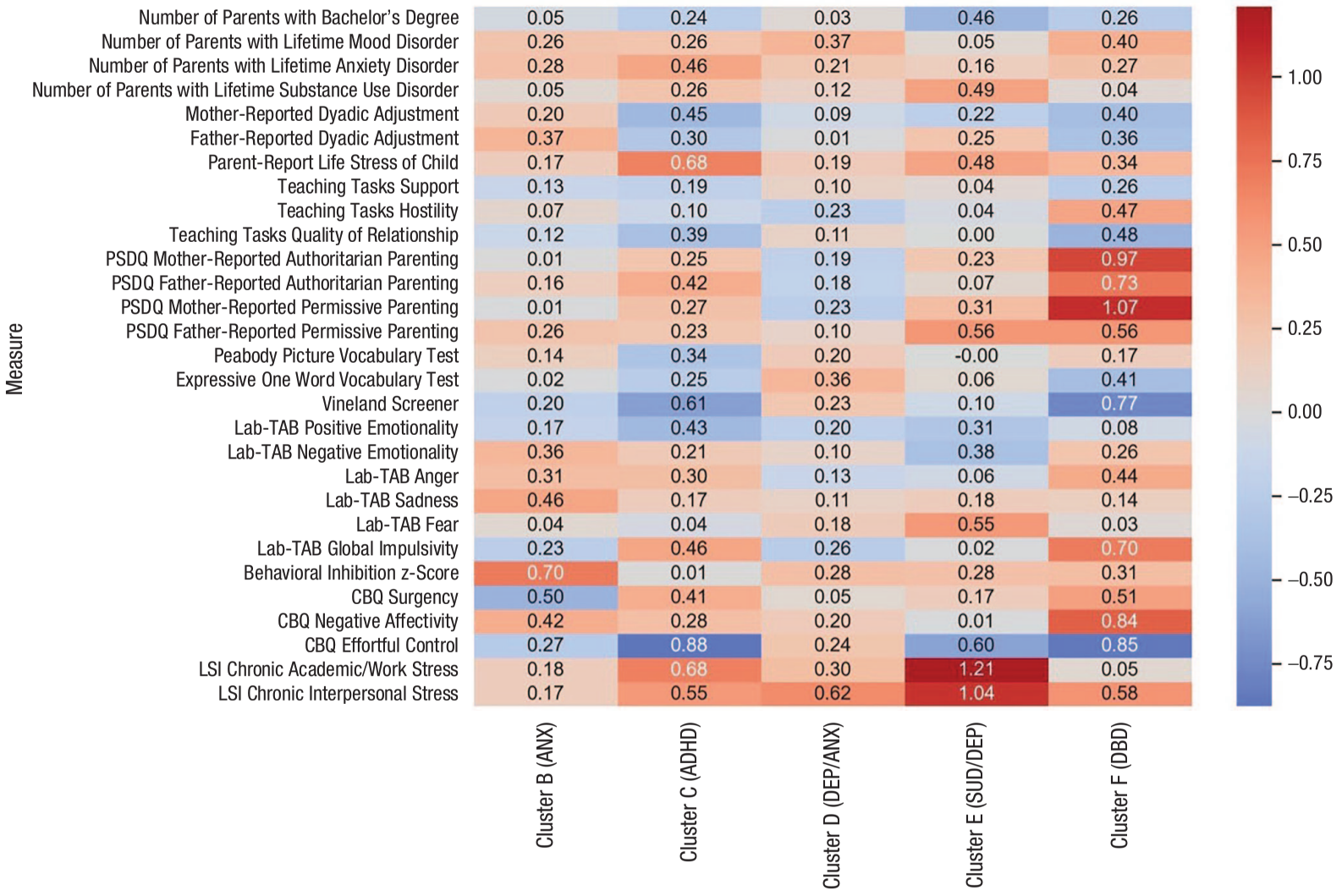


**Fig. 2.** Heat map of effect sizes for post hoc comparisons of all psychopathology clusters with Cluster A. Red cells indicate higher than Cluster A, and blue cells indicate lower than Cluster A. Effect sizes are expressed using Cohen's *d* (small = 0.2–0.5, medium = 0.5–0.8, large = 0.8–1.20). DEP = any depressive disorder, including not otherwise specified; ANX = any anxiety disorder, including not otherwise specified; ADHD = attention-deficit/hyperactivity disorder; DBD = disruptive behavior disorder; SUD = substance use disorder; PSDQ = Parenting Styles and Dimensions Questionnaire; Lab-TAB = Laboratory Temperament Assessment Battery; CBQ = Children's Behavior Questionnaire; LSI = Life Stress Interview.

Cluster E (SUD/DEP) had parents with significantly lower education and higher rates of SUD. Individuals in Cluster E also experienced higher levels of life stressors, higher father-reported permissive parenting, greater fear during the Lab-TAB, and lower parent-reported effortful control.

Cluster F (DBD group) had greater observed parental hostility and a poorer quality of parent-child relationship during the Teaching Tasks and higher mother- and father-reported authoritarian parenting and mother- and father-reported permissive parenting. The cluster also had significantly lower expressive-vocabulary and social-competence scores. They also displayed significantly higher levels of observed anger and global impulsivity. This cluster had significantly higher parent-reported surgency and negative affect and lower effortful control.

### *Age 18 functional outcomes*

Cluster B (ANX) did not differ from Cluster A (healthy) on academic/work and interpersonal functioning at age 18. In contrast, Cluster C (ADHD) displayed poorer functioning in both the academic/work and interpersonal domains compared with Cluster A. Cluster D (DEP/ANX) exhibited greater problems in interpersonal functioning. Cluster E (SUD/DEP) experienced significantly poorer academic/work and interpersonal functioning. Cluster F (DBD) displayed greater impairment in interpersonal functioning.

## Discussion

The importance of development, continuity, and course has long been emphasized in theory and research in psychopathology and developmental psychopathology (Bromet, 2015; Cicchetti & Rogosch, 2002) and is considered one of the criteria for judging the validity of diagnostic constructs (Robins & Guze, 1970). However, the development and course of psychopathology are seldom explicitly incorporated into classification systems (Klein, 2015; Tackett & Hallquist, 2022). To our knowledge, this study is the first to apply a longitudinal data-driven clustering algorithm to the range of common mental disorders. We applied this novel approach to rigorously collected data on six occasions from early childhood through late adolescence. The best-fitting solution revealed six clusters with distinct patterns of psychopathology at each time point and unique patterns of homotypic and heterotypic continuity across development.

Results of the present study are consistent with prior work that has reported both homotypic and heterotypic continuity of common mental disorders through adolescence and has shown greater continuity among, rather than between, internalizing and externalizing psychopathology (Caspi et al., 2020; Copeland et al., 2013; Finsaas et al., 2018; Healy et al. 2022; Oldehinkel & Ormel, 2023). However, we also found evidence of subgroups of individuals with unique patterns of sequential comorbidity (e.g., childhood anxiety that diminished with age vs. childhood anxiety that persisted into adolescence and the emergence of adolescent-onset depressive disorders vs. adolescent-onset depression and SUDs). Moreover, we found that the six clusters revealed in our analyses generally differed regarding sex distribution, early childhood predictors (i.e., parental education and psychopathology, early environment, temperament, cognitive and social functioning), and late-adolescent outcomes (i.e., functional impairment). Thus, this approach accounted for comorbidity and change in symptom presentation, creating more homogeneous subgroups of transdiagnostic psychopathology that were associated with different antecedents and outcomes.

We note, however, that the optimal cluster solution is derived from a particular prospective longitudinal data set and that this exact solution may not be replicated in other data sets using different samples, age groups, and measures. For example, the sample's age influences the prevalence of particular disorders (e.g., one would not expect to find much substance use before adolescence). In addition, the prevalence of disorders will differ depending on whether the sample is selected from the community or a clinical setting. Our community sample has very low rates of psychosis and eating disorders, which might be more common in a clinically referred sample. Thus, we regard this more as a proof of principle rather than as a definitive classification. Specifically, we argue that longitudinal clustering of psychopathology affords unique insights into patterns of continuity, sequential comorbidity, and developmental patterning not otherwise considered by cross-sectional classification systems (Lahey et al., 2014).

Cluster B (ANX) can be thought of as reflecting fear-related internalizing psychopathology, whereas Cluster D (DEP/ANX) reflects distress-related internalizing psychopathology. These clusters differ in that compared with Cluster A, Cluster B displayed higher rates of anxiety disorders at earlier ages and was characterized by early temperamental negative affect and behavioral inhibition according to both observational and parent-report measures. However, Cluster B did not exhibit notable comorbidity, and rates of anxiety disorders declined in adolescence, suggesting that this group is the healthiest of the psychopathology clusters. Indeed, this cluster did not differ from Cluster A (healthy) on parent psychopathology and parenting or functional

outcomes at age 18. Consistent with this, the five-factor solution did not produce a cluster like Cluster B; most participants in Cluster B were included in the healthy cluster in the five-factor solution.

Cluster D, by contrast, displayed gradually rising rates of anxiety disorders throughout adolescence and rates of depressive disorders that increased sharply in mid-late adolescence, suggesting a later onset but possibly more pernicious course of psychopathology than Cluster B. Cluster D had higher early expressive vocabulary, a higher rate of parental mood disorders, and poorer functional outcomes than Cluster A. When directly comparing Cluster B and Cluster D (see Table S1 in the Supplemental Material), we found that Cluster D had lower father-reported dyadic adjustment, higher parent-reported social competence, lower observed and parent-reported behavioral inhibition, and lower observed anger. Cluster D also had higher mother-reported surgency, higher mother-reported effortful control, and greater impairment in interpersonal functioning at age 18. Thus, Cluster D has generally better functioning that declines in adolescence, and Cluster B has worse functioning in childhood that improves in adolescence. These differences further strengthen the distinction between two clusters despite their diagnostic overlap.

Cluster E (DEP/SUD) was also characterized by increasing depression but was quite different from Cluster D (DEP/ANX) in that it was additionally marked by SUDs rather than anxiety disorders. Thus, it reflects a mixed internalizing/externalizing presentation. Depression and SUDs are frequently comorbid (Swendsen & Merikangas, 2000), and it is posited that individuals with depression might use substances as a form of self-medication or an attempt to cope with negative mood or stressful life events (Magee & Connell, 2021). Alternatively, substances can lead to depression through their biological effects or indirectly via functional consequences (Boden & Fergusson, 2011). Indeed, in this study, we found that Cluster E had the most significant associations with impaired academic and interpersonal functioning compared with Cluster A (healthy). Compared with Cluster A, Cluster E was also associated with higher levels of stress and temperamental fear in early life despite low rates of psychopathology at earlier ages that did not intensify until age 12. This cluster might reflect the internalizing (as opposed to externalizing) pathway to SUDs as discussed by Chassin et al. (2013). Note that although Cluster D was associated with a higher rate of parental depression than Cluster A, Cluster E had a higher rate of parental SUD than Cluster A, suggesting some specificity in familial etiological influences. This, using Cluster A as the reference group, Clusters D and E were distinguished by several early childhood risk factors despite both being characterized by increasing rates of depression in adolescence. When Clusters D and E were compared directly (see Table S3 in the Supplemental Material), we found that Cluster E was characterized by higher mother-reported permissive parenting, lower mother-reported effortful control, and greater academic impairment at age 18. Again, these differences underscore the utility of our longitudinal clustering approach in distinguishing phenotypes that can appear similar at single points in time but are very different from the perspective of risk factors and developmental course.

Clusters C (ADHD) and F (DBD) were the two externalizing clusters. Cluster C exhibited low levels of psychopathology in early childhood that intensified through later childhood and adolescence, manifesting mainly as ADHD with some co-occurring anxiety and depressive disorders and DBDs. Cluster F, by contrast, was characterized by high rates of DBDs in early childhood that declined but showed a small increase again in adolescence. Individuals in Cluster F also had modest but elevated rates of other disorders over the course of development. Despite the difference in the nature of their symptoms and the onset and course of psychopathology, compared with Cluster A (healthy) at age 3, Clusters C and F both displayed a poorer observed quality of parent-child relationship, higher levels of father-reported authoritarian parenting, lower social competence, greater observed impulsivity, lower parent-reported effortful control, greater parent-reported surgency, and poorer verbal abilities (although they differed on whether the problems were receptive or expressive). In addition, Cluster C was marked by a history of anxiety disorders in parents, consistent with some prior literature that has found familial coaggregation of ADHD and anxiety (Jarrett et al., 2016). Cluster F, by contrast, was associated with more problematic parenting styles and temperamental anger in early childhood compared with Cluster A, consistent with the high rate of DBDs in this cluster. When comparing them directly (see Table S2 in the Supplemental Material), we found that Clusters C and F differed in that Cluster C had lower expressive-vocabulary abilities at age 18 and Cluster F had higher mother-reported authoritarian and permissive parenting and mother-reported negative affect. These differences perhaps point to Cluster C having greater verbal-learning difficulties, whereas Cluster F was marked by greater mood dysregulation in early childhood. Regarding functional impairment, Cluster C displayed poorer academic functioning at age 18 compared with both Cluster A and Cluster F. This makes sense given the more problematic executive and language functioning in this cluster and the fact that their high rates of psychopathology persisted into late adolescence. Cluster F, on the other hand, did not show

later functional impairment, and it appears that their DBDs had largely remitted by late adolescence. In our community sample, we perhaps lack sufficient power to detect a small subgroup with early behavior problems and poor functional outcomes that corresponds to Moffitt's (2001) life-course-persistent subtype of conduct problems. Indeed, a seven-cluster solution produced clusters that were similar to Moffitt's typology, but the sample sizes were small, and overall model fit was poorer than for the six-cluster solution.

Overall, the distinction among the derived clusters was supported by a variety of early childhood risk factors and late-adolescent functional outcomes. This study is unique in its explicit consideration of developmentally based transdiagnostic subgroups with meaningful clinical, family, and temperamental correlates. The use of longitudinal *k*-means clustering to psychiatric diagnostic data is particularly novel and has utility for future research using syndrome- or symptom-level characterization of psychopathology. Although our clusters align for the most part with the existing literature, they provide a richer and more detailed perspective, illustrating different patterns of the development of psychopathology and the unfolding of sequential comorbidity and heterotypic continuity over distinct developmental periods.

However, the present study had several limitations. First, the sample was predominantly White and non-Hispanic and cannot be assumed to generalize to other ethnoracial groups because the relative homogeneity of the sample precluded an adequately powered analysis of racial and ethnic differences between clusters. Given evidence that rates of psychopathology and many clinical features (e.g., age of onset, severity, course, comorbid conditions) differ by sociodemographic characteristics, greater attention to sampling strategies and ethnoracial diversity is needed to understand how developmental patterning of psychopathology differs across groups (Wilson, 2024). Second, we assessed psychopathology every 3 years. Triennial assessments may have limited precision, particularly early in development, when change is rapid. Thus, we may have missed some onsets and offsets of episodes of psychopathology, particularly during the age-3 and age-6 assessments, when we focused on the 3 months before the interview instead of assessing the entire interval, as we did in subsequent waves. Third, some of the measures (i.e., DBD at age 9, rating of parent confidence in the Teaching Tasks) had interrater reliabilities in the upper end of the fair range, based on Shrout's (1998) recommendations. Lower interrater reliability could affect the prevalence of disorders and attenuate prediction by age-3 variables. However, almost all measures had reliabilities that were in the moderate-substantial range. Fourth, the severity of psychopathology was not considered. Clusters were derived based on the presence or absence of diagnoses at each time point. The incorporation of dimensional measures of psychopathology might provide additional nuance or even more distinctive groups. Future research using a dimensional approach to capture variability in symptom trajectories is warranted. Fifth, although this study emphasizes person-centered trajectories to understanding the patterning of homotypic and heterotypic continuity, the interpretation of each cluster is based on the proportion of individuals in that cluster with the specific diagnosis at that time point. Thus, some heterogeneity remains in each cluster because cluster members do not exhibit a uniform developmental course. Finally, the validation measures were assessed at either the age-3 or age-18 waves, two time points that were included in the clustering algorithm. Hence, they are not entirely independent of the clusters. This likely inflated some of the associations. For example, effect sizes of the association between maladaptive parenting and Cluster F assignment were extremely large, very possibly because the parenting measures were derived at age 3 and this cluster was characterized by an elevated rate of disruptive behavior disorders at age 3. Thus, the present cluster solution is offered as illustrative rather than definitive and provides a proof of principle that we hope will stimulate further work in longitudinal classification.

In conclusion, the results of the present study suggest that a data-driven person-centered approach to classifying psychopathology that includes development and course over time may be useful in identifying meaningful transdiagnostic groups. These groups differed on preschool assessments of parental education, family history of psychopathology, parenting and life stress, preschool temperament and cognitive and social development, and functional impairment at 18 years of age. The consideration of development and course in psychiatric classification may be crucial to address issues of heterogeneity and comorbidity and account for the patterning of homotypic and heterotypic continuity (Tackett & Hallquist, 2022).

## Transparency

*Action Editor:* Aleksandra Kaurin
*Editor:* Jennifer L. Tackett
*Author Contributions*

**Connor Lawhead:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing.
**Jamilah Silver:** Data curation; Writing – review & editing.
**Thomas M. Olino:** Resources; Supervision; Writing – review & editing.

**Loïc Labache:** Formal analysis; Writing – review & editing.
**Swanie Juhng:** Conceptualization; Investigation.
**H. Andrew Schwartz:** Conceptualization; Methodology; Writing – review & editing.
**Daniel N. Klein:** Conceptualization; Data curation; Funding acquisition; Investigation; Project administration; Supervision; Writing – review & editing.

*Declaration of Conflicting Interests*

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

*Funding*

Support for this research was provided through National Institute of Mental Health R01 MH069942 (D. N. Klein).

## ORCID iDs

Connor Lawhead https://orcid.org/0009-0005-2099-4844
Jamilah Silver https://orcid.org/0000-0002-8750-2876
Thomas M. Olino https://orcid.org/0000-0001-5139-8571
Loïc Labache https://orcid.org/0000-0002-5733-0743
Daniel N. Klein https://orcid.org/0000-0003-4582-6669

## Acknowledgments

We assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975. We obtained written consent from subjects.

## Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/21677026251357589

## Note

1. Some of the data from the present study have been used in other publications (e.g., Bufferd et al., 2012; Finsaas et al., 2018; Olino et al., 2010); however, no prior study has used these data to address the aims of the current study—that is, no researchers have applied factor, cluster, or latent-class/profile analysis to the diagnostic data, and we have not used a longitudinal clustering approach such as the one in the present study in any previous articles.

## References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. N. (2015). *Patterns of attachment: A psychological study of the strange situation*. Psychology press.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Angold, A., & Egger, H. L. (2007). Preschool psychopathology: Lessons for the lifespan. *Journal of Child Psychology & Psychiatry*, *48*(10), 961–966.

Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*, *74*(6), 1899–1917.

Boden, J. M., & Fergusson, D. M. (2011). Alcohol and depression. *Addiction*, *106*(5), 906–914.

Bromet, E. (2015). *Long term outcomes in psychopathology research*. Oxford University Press.

Brownell, R. (Ed.). (2000). *Expressive one-word picture vocabulary test: Manual*. Academic Therapy Publications.

Bufferd, S. J., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2011). Parent-reported mental health in preschoolers: Findings using a diagnostic interview. *Comprehensive Psychiatry*, *52*(4), 359–369. https://doi.org/10.1016/j.comppsych.2010.08.006

Bufferd, S. J., Dougherty, L. R., Carlson, G. A., Rose, S., & Klein, D. N. (2012). Psychiatric disorders in preschoolers: Continuity from ages 3 to 6. *American Journal of Psychiatry*, *169*(11), 1157–1164.

Bufferd, S. J., Dyson, M. W., Hernandez, I. G., & Wakschlag, L. S. (2016). Explicating the "developmental" in preschool psychopathology. In D. Cicchetti (Ed.), *Developmental psychopathology: Maladaptation and psychopathology* (3rd ed., pp. 152–186). John Wiley & Sons.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, *3*(1), 1–27.

Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., Harrington, H., Hogan, S., Poulton, R., Ramrakha, S., Rasmussen, L. J. H., Reuben, A., Richmond-Rakerd, L., Sugden, K., Wertz, J., Williams, B. S., & Moffitt, T. E. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. *JAMA Network Open*, *3*(4), Article e203221. https://doi.org/10.1001/jamanetworkopen.2020.3221

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

Chassin, L., Sher, K. J., Hussong, A., & Curran, P. (2013). The developmental psychopathology of alcohol use and alcohol disorders: Research achievements and future directions. *Development and Psychopathology*, *25*(4, Pt. 2), 1567–1584.

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, *70*(1), 6–20. https://doi.org/10.1037//0022-006x.70.1.6

Copeland, W. E., Adair, C. E., Smetanin, P., Stiff, D., Briante, C., Colman, I., Fergusson, D., Horwood, J., Poulton, R., Costello, E. J., & Angold, A. (2013). Diagnostic transitions from childhood to adolescence to early adulthood. *Journal of Child Psychology and Psychiatry*, *54*(7), 791–799.

Costello, E. J., Angold, A., March, J., & Fairbank, J. (1998). Life events and post-traumatic stress: The development of a new measure for children and adolescents. *Psychological Medicine*, *28*(6), 1275–1288.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–third edition (PPVT-III)* [Database record]. APA PsycTests.

Egeland, B., & Hiester, M. (1995). The long-term consequences of infant day-care and mother-infant attachment. *Child Development*, *66*(2), 474–485.

Egger, H. L., Ascher, B. H., & Angold, A. (1999). *Preschool age psychiatric assessment (PAPA)*. Duke University Medical Center.

Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B. K., & Angold, A. (2006). Test-retest reliability of the preschool age psychiatric assessment (PAPA). *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(5), 538–549.

Finsaas, M. C., Bufferd, S. M., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2018). Preschool psychiatric disorders: Homotypic and heterotypic continuity through middle childhood and early adolescence. *Psychological Medicine*, *48*, 2159–2168.

First, M. B., & Gibbon, M. (2004). The structured clinical interview for DSM-IV Axis I disorders (SCID-I) and the structured clinical interview for DSM-IV Axis II disorders (SCID-II). In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment* (pp. 134–143). John Wiley & Sons, Inc.

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment*, *23*(2), 337–353.

Genolini, C., Alacoque, X., Sentenac, M., & Arnaud, C. (2015). Kml and kml3d: R packages to cluster longitudinal data. *Journal of Statistical Software*, *65*, 1–34.

Genolini, C., Pingault, J. B., Driss, T., Cote, S., Tremblay, R. E., Vitaro, F., Arnaud, C., & Falissard, B. (2013). KmL3D: A non-parametric algorithm for clustering joint trajectories. *Computer Methods and Programs in Biomedicine*, *109*(1), 104–111.

Hammen, C., Adrian, C., Gordon, D., Burge, D., Jaenicke, C., & Hiroto, D. (1987). Children of depressed mothers: Maternal strain and symptom predictors of dysfunction. *Journal of Abnormal Psychology*, *96*(3), 190–198.

Harkness, K. L., & Monroe, S. M. (2016). The assessment and measurement of adult life stress: Basic premises, operational principles, and design requirements. *Journal of Abnormal Psychology*, *125*(5), 727–745.

Healy, C., Brannigan, R., Dooley, N., Staines, L., Keeley, H., Whelan, R., Clarke, M., Zammit, S., Kelleher, I., & Cannon, M. (2022). Person-centered trajectories of psychopathology from early childhood to late adolescence. *JAMA Network Open*, *5*(5), Article e229601. https://doi.org/10.1001/jamanetworkopen.2022.9601

Jarrett, M. A., Wolff, J. C., Davis, T. E., 3rd, Cowart, M. J., & Ollendick, T. H. (2016). Characteristics of children with ADHD and comorbid anxiety. *Journal of Attention Disorders*, *20*(7), 636–644.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988.

Kendler, K. S., Ohlsson, H., Sundquist, J., & Sundquist, K. (2023). Relationship of family genetic risk score with diagnostic trajectory in a Swedish National Sample of incident cases of major depression, bipolar disorder, other nonaffective psychosis, and schizophrenia. *JAMA Psychiatry*, *80*(3), 241–249.

Klein, D. N. (2015). Can course help reduce the heterogeneity of depressive disorders? In E. J. Bromet (Ed.), *Long-term outcomes in psychopathology research: Rethinking the scientific agenda* (pp. 32–52). Oxford University Press.

Klein, D. N., & Finsaas, M. C. (2017). The Stony Brook Temperament Study: Early antecedents and pathways to emotional disorders. *Child Development Perspectives*, *11*(4), 257–263.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477.

Kraepelin, E. (1919). *Dementia praecox and paraphrenia* (R. M. Barclay, Trans.). E. & S. Livingstone.

Lahey, B. B., Zald, D. H., Hakes, J. K., Krueger, R. F., & Rathouz, P. J. (2014). Patterns of heterotypic continuity associated with the cross-sectional correlational structure of prevalent mental disorders in adults. *JAMA Psychiatry*, *71*(9), 989–996.

Luby, J. L., Heffelfinger, A. K., Mrakotsky, C., Hessler, M. J., Brown, K. M., & Hildebrand, T. (2002). Preschool major depressive disorder: Preliminary validation for developmentally modified DSM-IV criteria. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(8), 928–937.

Lyneham, H. J., & Rapee, R. M. (2005). Agreement between telephone and in-person delivery of a structured interview for anxiety disorders in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(3), 274–282.

Magee, K. E., & Connell, A. M. (2021). The role of substance use coping in linking depression and alcohol use from late adolescence through early adulthood. *Experimental and Clinical Psychopharmacology*, *29*(6), 659–669.

Martinek, T., Jarczok, M., Rottler, E., Hartmann, A., Zeeck, A., Weiß, H., & von Wietersheim, J. (2023). Typical disease courses of patients with unipolar depressive disorder after in-patient treatments–Results of a cluster analysis of the INDDEP project. *Frontiers in Psychiatry*, *14*, Article 1081474. https://doi.org/10.3389/fpsyt.2023.1081474

Maughan, B., & Collishaw, S. (2015). Development and psychopathology: A life course perspective. In A. Thapar, D. S. Pine, J. F. Leckman, S. Scott, M. J. Snowling, & E. Taylor (Eds.), *Rutter's child and adolescent psychiatry* (pp. 1–16). John Wiley & Sons.

McGorry, P. D., Hickie, I. B., Yung, A. R., Pantelis, C., & Jackson, H. J. (2006). Clinical staging of psychiatric disorders: A heuristic framework for choosing earlier, safer and more effective interventions. *Australian & New Zealand Journal of Psychiatry*, *40*(8), 616–622.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159–179.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, *13*(2), 355–375.

Oldehinkel, A. J., & Ormel, J. (2023). Annual research review: Stability of psychopathology: Lessons learned from longitudinal population surveys. *Journal of Child Psychology and Psychiatry*, *64*(4), 489–502.

Olino, T. M., Klein, D. N., Dyson, M. W., Rose, S. A., & Durbin, C. E. (2010). Temperamental emotionality in preschool-aged children and depressive disorders in parents: Associations in a large community sample. *Journal of Abnormal Psychology*, *119*(3), 468–478.

Pfeifer, M., Goldsmith, H. H., Davidson, R. J., & Rickman, M. (2002). Continuity and change in inhibited and uninhibited children. *Child Development*, *73*(5), 1474–1485.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry*, *126*(7), 983–987.

Robinson, C. C., Mandleco, B., Olsen, S. F., & Hart, C. H. (1995). Authoritative, authoritarian, and permissive parenting practices - Development of a new measure. *Psychological Reports*, *77*(3), 819–830.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1997). Comparability of telephone and face-to-face interviews in assessing Axis I and II disorders. *American Journal of Psychiatry*, *154*(11), 1593–1598.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*(5), 1394–1408.

Schulte, E. C., Kondofersky, I., Budde, M., Papiol, S., Senner, F., Schaupp, S. K., Reich-Erkelenz, D., Klohn-Saghatolislam, F., Kalman, J. L., Gade, K., Hake, M., Comes, A. L., Anderson-Schmidt, H., Adorjan, K., Juckel, G., Schmauss, M., Zimmermann, J., Reimer, J., Wiltfang, J., . . . Schulze, T. G. (2022). A novel longitudinal clustering approach to psychopathology across diagnostic entities in the hospital-based PsyCourse study. *Schizophrenia Research*, *244*, 29–38.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, *7*(3), 301–317.

Spanier, G. B. (1976). Measuring dyadic adjustment - New scales for assessing quality of marriage and similar dyads. *Journal of Marriage and the Family*, *38*(1), 15–28.

Sparrow, S. S., Carter, A. S., & Cicchetti, D. V. (1993). *Vineland Screener: Overview, reliability, validity, administration, and scoring*. Yale University Child Study Center.

Swendsen, J. D., & Merikangas, K. R. (2000). The comorbidity of depression and substance use disorders. *Clinical Psychology Review*, *20*(2), 173–189.

Tackett, J. L., & Hallquist, M. (2022). The need to grow: Developmental considerations and challenges for modern psychiatric taxonomies. *Journal of Psychopathology and Clinical Science*, *131*(6), 660–663.

Tian, C., Manfei, X. U., Justin, T. U., Hongyue, W., & Xiaohui, N. I. U. (2018). Relationship between omnibus and post-hoc tests: An investigation of performance of the F test in ANOVA. *Shanghai Archives of Psychiatry*, *30*(1), 60–64. https://doi.org/10.11919/j.issn.1002-0829.218014

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Wilson, S. (2024). Sociodemographic reporting and sample composition over 3 decades of psychopathology research: A systematic review and quantitative synthesis. *Journal of Psychopathology and Clinical Science*, *133*(1), 20–36. https://doi.org/10.1037/abn0000871