Individual Differences in the Stability and Change of Childhood Depression: A Growth Mixture Model With Structured Residuals

Matt Hawrilenko [iD]
University of Washington School of Medicine

Katherine E. Masyn [iD]
Georgia, State University

Janine Cerutti [iD]
Massachusetts General Hospital

Erin C. Dunn [iD]
Massachusetts General Hospital and Harvard Medical School and Center on the Developing Child at Harvard University

Studies of developmental trajectories of depression are important for understanding depression etiology. Existing studies have been limited by short time frames and no studies have explored a key factor: differential patterns of responding to life events. This article introduces a novel analytic technique, growth mixture modeling with structured residuals, to examine the course of youth depression in a large, prospective cohort ($N = 11,641$, ages 4–16.5, 96% White). Age-specific critical points were identified at ages 8 and 13 where depression symptoms spiked for a minority of children. Most depression risk was due to dynamic responses to environmental events, drawn not from a small pool of persistently depressed children, but a larger pool of children who varied across higher and lower symptom levels.

Depression is one of the most common, costly, and disabling mental disorders worldwide (James et al., 2018) with lifetime prevalence estimates of 11.7% among adolescents (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015) and 16.6% among adults in the United States (Kessler et al., 2005). A large proportion of depression cases begin early in development, with one-third of people who experience depression having their first onset before age 21 (Zisook et al., 2007). Youth who experience depression are a particularly vulnerable group, as they are at an increased risk of suicide (Gould, Greenberg, Velting, & Shaffer, 2003), substance abuse (Lai, Cleary, Sitharthan, & Hunt, 2015), and are more likely to experience recurrent episodes of depression as adults (Lewinsohn, Rohde, Seeley, Klein, & Gotlib, 2000). In fact, earlier onsets of depression are associated with worse illness course and outcome into adulthood (Zisook et al., 2007). These findings underscore the need to understand the etiology and course of depression over time in order to prevent and treat the disorder as early in the lifespan as possible.

Challenges in Studying Depression Trajectories

However, efforts to characterize the etiology and course of depression have had limited success due to depression's heterogenous nature (van Loo, de Jonge, Romeijn, Kessler, & Schoevers, 2012). Here, heterogeneity refers to the nature of depression itself as a phenomenon with a variable symptom profile and variable developmental course regulated by time-invariant and time-varying processes. There are at least three major challenges related to

depression heterogeneity that may lead researchers to arrive at erroneous, inconsistent, or conflicting findings regarding the etiology and course of depression.

The first major challenge is that depression is difficult to measure both in childhood and across the life course because no single definition of "depression" exists. According to the Diagnostic and Statistical Manual of Mental Disorders fifth edition (American Psychiatric Association, 2013), there are 227 different ways to meet the diagnostic threshold for a major depressive episode, owing to different combinations of affective, cognitive, and behavioral symptoms (Galatzer-Levy & Bryant, 2013). Most troublingly, depression phenotypes may differ by age because symptom type, frequency, and severity are more or less prominent at different ages (Garvey & Schaffer, 1994; Hegeman, Kok, van der Mast, & Giltay, 2012). Even with the *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.'s wide inclusion criteria, the boundaries for what constitutes "depressed" may be too narrow, as subclinical depressive symptoms have been associated with substantial functional impairment (Allen, Chango, Szwedo, & Schad, 2014). Thus, depression has been alternately viewed not just as a binary construct (where people either have or do not have depression), but as a dimensional construct (everyone has *some amount* of depressive symptoms), or a combination of the two (Kotov et al., 2017). Reflecting these nuances, the most widely used scales to measure depressive symptoms vary in the affective, cognitive, and behavioral symptoms measured and consequently are only moderately correlated (Fried, 2017), meaning that a person's level of "depression" could be assessed very differently depending on the instrument used to measure it.

The second challenge is that depression emerges and changes across development; that is, it is an age-dependent trajectory of symptoms. Depression is rare in childhood until age 11, but then has larger yearly increases in prevalence through age 15 (Avenevoli et al., 2015). This age-dependence suggests that age-related events likely influence the onset of the disorder. Such events may span everything from biological changes including puberty (Graber, Lewinsohn, Seeley, & Brooks-Gunn, 1997) to sociocultural factors like psychosocial stress (Hyde, Mezulis, & Abramson, 2008). Moreover, developmental trajectories of depressive symptoms may be systematically different across children, meaning that different children may develop depression at different times via different pathways (Ellis et al., 2017; Shore, Toumbourou, Lewis, & Kremer, 2018).

The third challenge is that depression varies in how long symptoms last (chronicity) and how frequently symptoms return (recurrence). Some young people have minimal symptoms throughout their lives, whereas others have persistently high symptoms, and others have symptoms that fluctuate more dynamically across levels of severity (Hosenfeld et al., 2015; Lorenzo-Luaces, 2015). Prior studies in adolescents have found that the duration of a depressive episode can vary in length from 2 weeks to 10 years, and that one-third of recovered 14- to 18-year-olds had a recurrence within 4 years (Lewinsohn, Clarke, Seeley, & Rohde, 1994).

Renitent and Reversing Depression Symptomology

We posit that chronicity and recurrence may in part be driven by *renitent responding* and its opposite, *reversing responding* (collectively called *responding*). We define renitent responding as the degree to which a deviation from one's own average symptom levels perpetuates across time. That is, when a life event causes a person to become depressed, they remain depressed for long periods of time; when a life event causes them to feel better, this too perpetuates. Renitent responding is different from the concept of inertia (Kuppens et al., 2012) in that inertia refers to individuals "sticking" to a particular individual-specific mean course of symptomology while renitent responding deals with the persistence of deviations from the individual course due to exogenous factors (e.g., external life events). Individuals with renitent responding may have high or low chronicity in deviations from their unique course of symptomology depending on whether they experience adverse or positive events. In contrast to renitent responding, we define reversing responding as the degree to which a deviation from one's own average symptom levels reverses across time. That is, when a life event causes a person to become depressed, they feel better quickly; when a life event causes a person to feel better, their symptoms relapse quickly. People with reversing responding may transition more frequently from recovery to recurrence. Different patterns of renitent and reversing responding are shown in Figure 1. Some single-subject time-series research has shown that extreme levels of both positive *and* negative autocorrelations in depression scores over time are the characteristic of individuals with higher depression levels on average (Houben, Van Den Noortgate, & Kuppens, 2015). However, it remains unclear whether this is due to systematic inertia (depression causes itself), idiosyncratic

deviations from individual mean course due to exogenous factors such as independent life events, or both.

As posited by developmental psychopathology theories (Curran & Bauer, 2011), patterns of responding during any given depressive episode may be regulated by personal characteristics. For example, young adults with depression have been shown to employ coping behaviors that make their depressive symptoms better or worse (Bolger, 1990; Compas, Orosan, & Grant, 1993). Such behaviors, including reaching out for social support or withdrawing from social interactions, can shape responding during a depressive episode. Vulnerability characteristics, including early childhood abuse, negative life events, parental psychopathology, and chronic physical disorders (Galatzer-Levy & Bryant, 2013; Ten Have et al., 2018), may increase people's tendency for renitent responding initiated by a deviation of increased symptomology, whereas sleep, healthy diet, social support, and positive coping (Cairns, Yap, Pilkington, & Jorm, 2014) may lead to a balanced, middle ground between renitent and reversing responding initiated by a deviation of decreased symptomology. Unpacking how personal characteristics and environmental experiences shape renitent and reversing responding may be a key component to understanding depression heterogeneity (Franklin, Jamieson, Glenn, & Nock, 2015).

Modeling Heterogeneity in Depression Trajectories

Growth mixture modeling (GMM; Muthen & Muthen, 2000) has emerged as an analytic technique to characterize heterogeneity in developmental processes. In brief, GMM seeks to capture person-centered differences using a finite set of subpopulations (i.e., mixtures) with qualitatively distinct mean trajectories or group-specific patterns of systematic change in a developmental outcome over time. It is a *latent* variable approach because subgroup membership is not directly observed, but rather indirectly indicated by average symptom levels and change patterns over time. A major advantage of GMM over standard latent growth curve modeling is that it relaxes the assumption that all individuals are drawn from a single homogeneous population, so inferences can be made within and across more meaningful subgroups (e.g., different risk factors can be explored for people with minimal symptoms vs. those who develop depression at younger ages vs. those who develop depression at older ages). In their recent review of the literature, Ellis et al. (2017) found 18 studies that examined depressive symptom trajectories in children and adolescents. The studies are difficult to parsimoniously summarize because findings on the numbers of classes and the development of depression over time varied according to the age ranges included, and whether community or clinical samples were used. However, despite methodological differences, all studies found support for multiple classes over a single class, and nearly all studies identified a group with persistently minimal symptoms, as well as a group with symptoms that increase over time.

GMM has also been used to model developmental changes in internalizing symptoms, a closely related construct to depression. Internalizing symptoms consist of negative internal experiences such
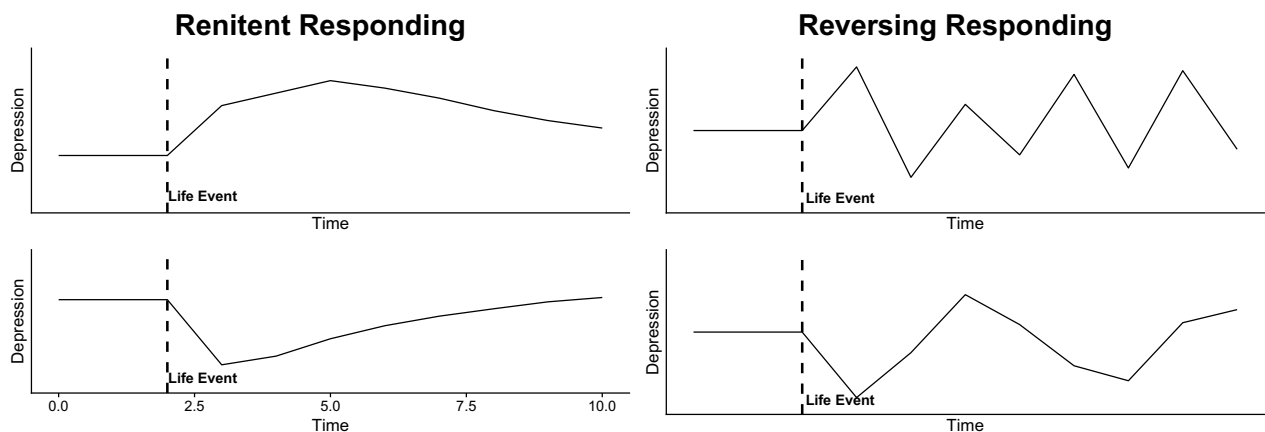

Figure 1. People may respond either to negative (top) or positive (bottom) life events. The average amount of time before people pass back through their own mean may vary across people.

as anxiety, sadness, and somatic issues. These GMM studies have found that symptom trajectories are best characterized by four to six subgroups (Dekker et al., 2007; Edwards et al., 2014), capturing children with consistently low symptoms, persistently high symptoms, low symptoms that increased during the adolescent years, and high symptoms that somewhat decreased as children aged.

For example, using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), Edwards et al. (2014) examined trajectories of internalizing symptoms from ages 4 through 11.5. They found that 75% of the sample had low symptoms across childhood, with the rest being divided into four subgroups consistent with previous analyses: a group with persistently high symptoms, a group whose symptoms increased across childhood, a group whose symptoms decreased across childhood, and a group whose symptoms rose to a moderate symptom level around age 8 and then tapered off by age 11. In another GMM analysis of ALSPAC, Rice et al. (2019) used latent class growth analysis, a special form of GMM that assumes there is no within-class variability in growth factors, and similar to Edwards and colleagues, they characterized 74% of people as having persistently low risk, and the remainder as having early adolescent onset or late adolescent onset depressive symptoms.

Together, GMM studies of depression and internalizing symptoms demonstrate that subtyping depression by its developmental course can paint a clearer picture of qualitatively different courses of depressive symptoms, which can aid in generating hypotheses about their etiological determinants. Despite their utility, existing GMM studies of depression have been limited in four primary ways.

First, all studies ignored the first major heterogeneity challenge described previously: the problem of measurement. These studies used only a single observed scale scored based on an aggregation of responses (typically a sum or average) to a single instrument over time. This type of score calculation makes the rather untenable assumption that the symptoms all function equally well as measures of underlying depression at each time point and that the items function the same across time, that is, at different ages. Given the heterogeneity in both symptoms and measures of depression, better measuring depressive symptoms using multiple measures, and testing and accommodating how the measurement model changes over time, can strengthen conclusions and lead to more replicable results.

Second, existing studies are limited because they mostly focus on a narrow range of ages, with only a few including children younger than 10 years old (Dekker et al., 2007; Edwards et al., 2014; Whalen et al., 2016). However, developmental pathways to depression likely begin much earlier than adolescence, as certain depressive symptoms have been shown to manifest as early as age 5 (Whalen, Sylvester, & Luby, 2017). Furthermore, childhood studies of depressive symptom trajectories have followed participants for an average of 3–6 years (Ellis et al., 2017), limiting the opportunity to track longer term developmental patterns. Thus, the extent to which developmental profiles of depression symptoms vary from childhood through adolescence remains unknown.

Third, with some recent exceptions (see review by Musliner, Munk-Olsen, Eaton, & Zandi, 2016), little work has linked these developmental trajectories to their etiological determinants. Linking discrete developmental trajectory typologies to the social and biological factors that set these trajectories in motion is critical for understanding how risk factors differentially influence the development of depression and can lead to identifying, as early on as possible, children most at-risk.

Finally, no existing studies have addressed the challenge that heterogeneity in the change and stability in depression symptomology may include two systematic components: One that is specific to each person (between-person variability in the underlying, individually varying mean levels and change as a function of age) and one that is person specific and time-anchored (within-person, time-specific deviations from the underlying, person-specific trajectory). This is a limitation GMM shares with other more conventional multilevel models and structural equation models for repeated measures (e.g., latent growth curve modeling). Although standard GMMs attempt to identify subgroups of individuals with homogeneous patterns of change, they have been criticized because even when GMMs ostensibly fit the data well, it does not guarantee that the subgroup mean trajectories generalize to any individual within the latent class (Bauer, 2007), potentially leading to false conclusions about developmental processes. One possible reason for discrepancies between group and individual level processes is the presence of within-person processes that are not explicitly modeled in standard GMMs. Because all existing studies have developed latent classes as a smooth function of age alone, they have ignored systematic, time-anchored, within-person processes related to stability

and change, such as renitent and reversing responding. Identifying systematic differences in patterns of responding may help identify groups of people who cope more or less effectively with the time-anchored life events that lead to their depression, and also elucidate malleable risk factors (Franklin et al., 2015).

The Growth Mixture Model With Structured Residuals

To partially address this last limitation, we introduce a novel form of GMM: the growth mixture model with structured residuals (GMM-SR). The GMM-SR combines the advantages of a growth model—known as the latent growth model with structured residuals (LGM-SR; Curran, Howard, Bainter, Lane, & McGinley, 2014)—with the advantages of GMM. In traditional latent growth models, depression scores are modeled only as a function of age, and a trajectory is calculated for each individual that describes their age-related change (Figure 2A). In the LGM-SR, an individual's depression scores are separated into age-related trajectories and time-anchored patterns of change (e.g., renitent responding and reversing responding), modeled in the form of a structured residual (Figure 2B). In this model, the residual is the difference between the observed depression score and the depression score expected based on each child's underlying (mean) growth trajectory. The time-specific residual represents the influence of all unmodeled variation in depression scores (e.g., exogenous life events) at a particular age. The *structured residual* imposes an autoregressive structure whereby each residual is systematically related to the one immediately before it and immediately after it. Thus, when an unmodeled life event causes a child's depression to go higher or lower than their own, underlying, age-based trajectory, the autoregressive term represents how the child's symptoms responded to that event. A positive autoregression represents renitent responding, or the tendency of a child's depression level to remain above or below their expected score from one time point to the next. A negative autoregression represents reversing responding, or the tendency of a child's depression level to systematically fluctuate across higher and lower scores. An autoregression of zero (the assumption of traditional growth models) means that the child's depression was well-modeled by the age-based trajectory alone and did not have any systematic, within-person dependencies over time due to unmodeled events.

The GMM-SR introduces the structured residual into the mixture model and thus allows for qualitatively distinct heterogeneity captured by the latent classes to be characterized by both class-specific mean growth trajectories (the GMM part) and time-anchored patterns of change and stability (the SR part). Figure 2 depicts hypothetical differences that might emerge between the GMM and GMM-SR, and includes examples where the standard GMM may incorrectly disaggregate people who should be grouped together (Figures 2C and 2D), and incorrectly aggregate people who should be in separate groups (Figures 2E and 2F). Note that the GMM-SR does not force such differences to emerge when they are better explained by a simpler model with only class-specific mean trajectories as a function of age. The standard GMM is nested within the GMM-SR and thus, if no patterning to depression scores were present after accounting for age-related changes, the structured residual autocorrelations would be zero, leaving a standard GMM. Structured residuals would be important, however, if the probability of remaining above or below one's own expected depression level were conditional on having been above or below one's expected level at the previous time point. In this regard, the GMM-SR may capture both age-dependent patterns of cycling through high and low symptoms as well as within-person, time-anchored, renitent and reversing responding. Determining whether a particular class represents a pattern of responding to environmental events versus a biologically determined cycle of high and low symptoms may be difficult, but can be aided by interpreting findings in concert with study design (i.e., which types of events or processes are likely to transact systematically over the time windows observed in the study) and by including covariates theoretically related to one specific pattern of responding (e.g., social and biological factors).

Current Study

The primary goal of this study was to characterize population heterogeneity in the person-specific stability and change of depression symptomology spanning across portions of the first two decades of life—from ages 4 through 16.5 years—using data from a unique ongoing longitudinal sample, the Avon Longitudinal Study of Parents and Children (ALSPAC). To simultaneously explore between-person differences in underlying age-based trajectories as well as between-person differences in the within-person, time-anchored stability and change in depressive
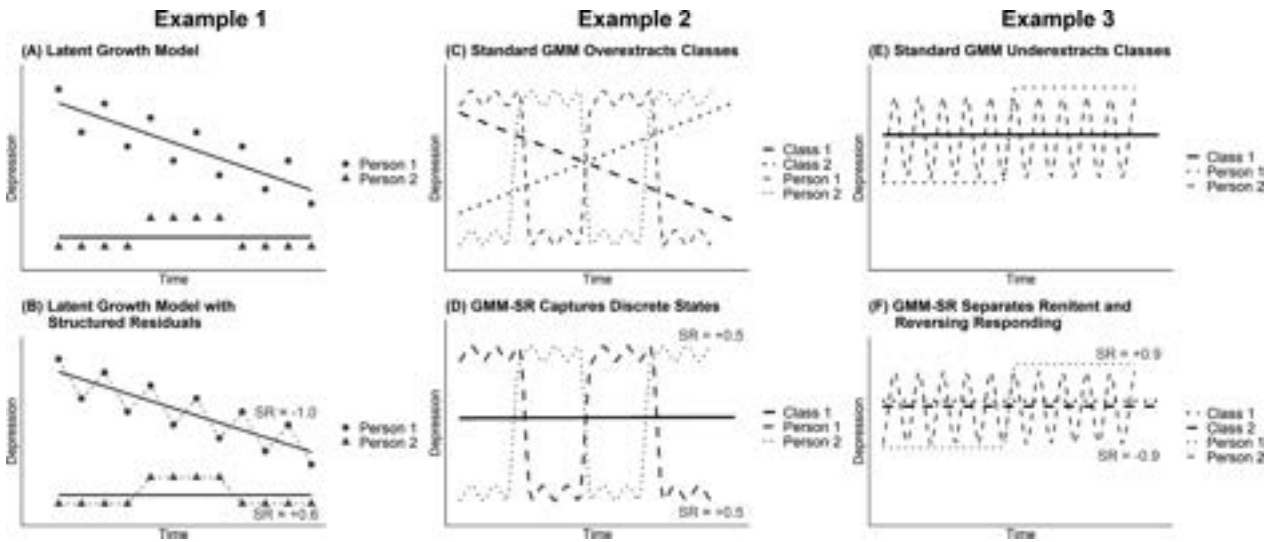
Figure 2. Example 1 shows the information captured by a latent growth model in the form of intercepts and trajectories (Panel A) and the additional information captured by the latent growth model with structured residuals in the form of positive and negative structured residual autocorrelations (Panel B). Example 2 shows people with identical symptom levels and time-anchored within-person symptom fluctuations who may be incorrectly separated into two different groups in a standard GMM that does not account for autocorrelation (Panel C), as opposed to a GMM-SR Panel D). Example 3 shows one person with reversing responding (SR = −0.9) who is changing rapidly, while the other with renitent responding (SR = +0.9) who is experiencing more sustained bouts of depression. These two people might be grouped into a single "high" class in a standard GMM (Panel E) but correctly separated into distinct classes in a GMM-SR (Panel F).

Note GMM = growth mixture model; SR = structured residuals.

symptoms relative to their underlying individual trajectories, we introduce the GMM-SR. A secondary goal of this study was to explore how five of the most common and impactful social factors related to depression influenced membership in emergent latent subgroups. We address the aforementioned limitations of previous GMM studies by modeling depression as a latent variable, measured by multiple instruments assessing depressive symptoms across time, and accounting for the possibility that symptoms functioned differently at different ages. We used this latent depression variable in a GMM-SR to characterize depression trajectories, including patterns of renitent responding and reversing responding, across this 13-year period. This was an exploratory study, as we did not test specific hypotheses about the number or nature of latent classes. By jointly characterizing both age-related change and patterns of responding and identifying etiological factors for both, our aim was to better describe heterogeneity in the etiology of depression.

Method

Sample and Procedures

ALSPAC sampled children born to mothers who were living in the county of Avon England (120 miles west of London) with estimated delivery dates between April 1991 and December 1992 (Fraser et al., 2012). When the oldest children were approximately 7 years old, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. 15,454 eligible pregnant women agreed to participate and were enrolled in the study, which resulted in 15,589 fetuses and a sample size of 14,901 who were alive at 1 year of age. The study Web site contains details of all the data that are available through a fully searchable data dictionary: www.bristol.ac.uk/alspac/researchers/our-data/. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee.

This study is based on an analytic sample of 11,641 children who had data on at least one measure of depressive symptoms between ages 4 and 16.5. Children were 51% male and came from households with a variety of SES backgrounds, with 27% of mothers having < 11 years of secondary education (i.e. O-level) and 14% having a degree or more. The sample was primarily White (95.7%). Nearly one in four children had mothers who met criteria for psychopathology between birth and 3 years old. Twelve percent of children lived in a household with a single parent for at least part of

their first 4 years, and 13% met criteria for neighborhood disadvantage. Just 6% of children experienced emotional or physical abuse at least once through 3.5 years of age. Children included in the analytic sample did not differ from those who were excluded with respect to sex. However, children who were excluded tended to be non-White (9.4% of excluded children vs. 4.3% of included children), have lower levels of maternal education (48% of excluded children had the lowest level of education vs. 27% of included children), and had mothers that were more likely to be unmarried (40% vs. 22%).

Measures

Outcome: Youth Depressive Symptoms

We used three measures that tapped depressive symptoms, all reported by mothers. These measures were derived from two subscales of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), and the Short Mood and Feelings Questionnaire (SMFQ; Angold et al., 1995). Questionnaires were selected, in part, because they tapped the most relevant depression symptoms at each age. The latent variable, discussed in the following section, is interpreted as the common cause of all symptoms across questionnaires.

Strengths and Difficulties Questionnaire. The SDQ was used to assess child emotional and behavioral problems. The SDQ is a dimensional rating of child psychopathology commonly used in epidemiology studies and has excellent psychometric properties (Muris, Meesters, & van den Berg, 2003). The SDQ has five subscales containing five items each and is rated on a three-point scale (0 = *not true*, 1 = *somewhat true*, or 2 = *certainly true*), capturing the child's behavior and feelings within the past 6 months. This study used two of the five SDQ subscales, child *emotional problems* and *peer difficulties*, which are often combined to represent "internalizing symptoms." Mothers completed the SDQ using mailed questionnaires at seven time points when their child was ages 4, 7, 8, 9.5, 11.5, 13, and 16.5 years.

Short Mood and Feelings Questionnaire. The SMFQ is a 13-item measure with three-point responses (*not true, sometimes, true*) often used in epidemiological studies to assess depressive symptoms in adolescents (Patton et al., 2008). The SMFQ has been shown in diverse populations to differentiate clinic from community samples and correlates highly with questionnaire and interview measures of psychopathology and clinician-rated diagnoses of

depression in early and late adolescence (Turner, Joinson, Peters, Wiles, & Lewis, 2014). Mothers completed the SMFQ by mailed questionnaire at three time points when their child was 11.5, 13, and 16.5 years.

Social Factors

We examined the role of five social factors as predictors of trajectory class membership: (a) maternal education (as a measure of socioeconomic position; four categories), (b) binary presence/absence of maternal psychopathology, (c) binary presence/absence of caregiver physical/emotional cruelty, (d) binary indicator of living in a single-parent household, and (e) binary indicator of neighborhood disadvantage (e.g., moderate levels of vandalism, worry about burglary, etc.). To ensure these factors preceded the measurement of depressive symptoms, all factors were measured at or before the age of 4 years. All social factors were chosen based on previous studies showing they preceded the development of psychopathology in other samples (Felitti et al., 1998; Petterson & Albers, 2001) and coded consistently with a prior ALSPAC study (Dunn et al., 2018; see Appendix S1 for further details).

Analytic Strategy

Analyses were conducted in four phases: (a) construction of a measurement model, (b) identification of a growth model reflecting the functional form of the change process, (c) latent class enumeration and selection of the unconditional growth mixture model, and (d) exploration of the predictors of latent trajectory class membership.

Latent Depression Score Measurement Model

A strength of confirmatory factor models is their ability to combine different subscales into a single score even when—as is the case here—not all scales are administered at all timepoints (Graham, Taylor, Olchowski, & Cumsille, 2006). Using confirmatory factor analysis, we combined the two SDQ subscales and SMFQ into a single score representing the cause of their shared variability, which we interpret as *depression*. Item level confirmatory factor models were used to test and account for the possibility that questionnaire items had different relationships to the latent construct, depression, at different ages (i.e., measurement invariance). Doing so is important to ensure that changes in the depression indicators over time are indicative of

change in the underlying latent depression construct, rather than measurement artifacts. Based on results from these preliminary analyses (see Appendix S2), which revealed that a number of items had age-related changes caused by factors other than depression (i.e., scalar noninvariance), parcels were constructed, meaning all items were divided into subsets and each subset was averaged (Little, Cunningham, Shahar, & Widaman, 2002). There were five parcels: the SDQ-emotions subscale, SDQ-peer difficulties subscale, and three groups of items from the SMFQ. Each parcel was constructed so that its items followed a similar pattern of scalar noninvariance. Equality constraints on noninvariant parcels were released across waves. Because incorporating the full item-level model in the mixture analysis was computationally intractable, we used this measurement model to generate empirical Bayes factor score estimates, which represent the latent depression score at each time point, and used these estimates in subsequent growth models and mixture models (Curran, Cole, Bauer, Rothenberg, & Hussong, 2018).

Growth Model

To examine the importance of patterns of responding, latent growth models were compared to the LGM-SR (Curran et al., 2014). We considered several functional forms of trajectories (i.e., linear through quintic polynomials) to capture curvilinear change over time and determined the best model using global and comparative fit statistics. In the LGM-SRs, we constrained the structured residual autoregression to be equal across time and used a total of three phantom variables between the longest time intervals to reduce the influence of unequally spaced measures on the autoregressive parameter (the longitudinal design is shown in Appendix S3). Phantom variables are a coding trick that impose constraints on the model by inserting missing values for all participants. For example, if depression were measured at times 1, 3, and 4, a model without phantom variables would assume that the autoregression from time 1 to time 3 ($AR_{1 \rightarrow 3}$) should equal the autoregression from time 3 to time 4 ($AR_{3 \rightarrow 4}$). Inserting a phantom variable at time 2 imposes the proper constraint that $AR_{1 \rightarrow 3} = AR_{1 \rightarrow 2} \times AR_{2 \rightarrow 3} = (AR_{3 \rightarrow 4})^2$.

Growth Mixture Model Class Enumeration

Growth mixture models using up to eight classes were fit to the data, with varying levels of constraints placed on the covariance structure. A wide variety of fit statistics were used to empirically determine the best model. The final model was chosen based on fit to the data (fit statistics), fit to the individual (visual inspection to ensure that latent classes accurately represented trajectories of the people within them), and the model's ability to provide theoretically useful distinctions across classes (Masyn, 2013).

Predictors of Latent Class Membership

The two-step method (Bakk & Kuha, 2018) was used to examine how the probability of latent class membership varied across social factors. The first of the "two-steps" refers to determining the best growth mixture model (as above). In the second step, the growth mixture portion of the model was held fixed, while class probabilities and the relationships between predictors and latent classes were freely estimated. To manage Type I error, multinomial contrasts for predictors were block-tested using a Wald Test with the false disocvery rate (Benjamini & Hochberg, 1995) set to 5%.

Missing Data

Of the 11,641 children included in the initial sample, missingness varied across follow-up waves from a low of 19% (age 4) to a high of 51% (age 16.5). Data missingness was handled in two stages. The purpose of the first stage was to impute information on depression scores that accounted for their changing measurement properties over time. We imputed empirical Bayes factor score estimates from the measurement model. To improve imputation quality, we included additional baseline characteristics related to missingness as saturated correlates (see Appendix S4; Collins, Schafer, & Kam, 2001). Thus, individuals providing data at even a single time point received factor score estimates at all time points, leveraging information from their available depression data and baseline covariates. An alternative to imputing factor scores in the measurement model would have been to impute at the item-level before calculating factor scores. However, the item-level imputation model yielded unstable estimates, perhaps due to the large number of ordinal response items. For this reason, we used the relatively simpler approach of relying on maximum likelihood estimation directly in the factor score model (Curran et al., 2018).

Because imputation models must explicitly include all relationships tested in the analytic

model, the purpose of the second stage of missingness handling was to include latent class membership in the covariate imputation model. Mixture model parameters were fixed to the estimates derived from the final mixture model and means and covariances of predictors were free to vary across classes (this procedure is analogous to the two-step method; Bakk & Kuha, 2018). We used multiple imputation for this second stage rather than full information maximum likelihood because the latter was computationally intractable given that there is essentially a latent class-varying latent factor for each missing covariate. Mplus Version 8.0 was used for all analyses (Muthén and Muthén, 1998–2017).

Results

Sample Characteristics

Descriptive statistics and item correlations for predictor variables and depression factor scores are presented in Table 1. Mean depression scores did not change much over time, although variability increased with age. Predictors had small positive correlations with each other and with depression, suggesting that social adversities represented distinct constructs, tended to co-occur, and were associated with higher levels of depression.

Measurement Model

We tested whether the underlying latent depression construct accounted for all changes over time in each parcel of questionnaire responses by constraining the intercepts of each parcel equal across waves (i.e., full scalar invariance). Fit of the full scalar invariance model was rejected when compared to the metric model, which allowed intercepts for each parcel to vary independently of one another. Misfit was driven by systematic drops in SDQ peer difficulties from ages 4 to 8, and systematic increases in several SMFQ items reflecting negative emotions at age 16.5, indicating these changes were not caused by latent depression. To ensure that factor score estimates captured only changes driven by the latent depression construct, we freed the intercept constraints at these waves for both parcels, thereby restricting the noninvariant parcels to contribute to information about the factor score within each wave, but not to changes in factor scores across waves. Fit of this partial scalar model was excellent (comparative fit index [CFI] = .999, Tucker–Lewis index [TLI] = .998, root mean square

error of approximation [RMSEA] = .015) and superior to the fit of the model with full scalar invariance ($\chi^2_{diff}(5) = 989.11$, $p < .0001$). Standardized loadings for parcels were high for SDQ-emotional difficulties, nearly as strong for each SMFQ parcel, and weakest for SDQ-peer difficulties (Appendix S2). Factor score distributions had moderate levels of skew and kurtosis. Because skew and kurtosis may cause GMMs to extract latent classes that characterize properties of the distribution rather than true population-wide heterogeneity (Bauer & Curran, 2003), factor scores were log-transformed, resulting in normal distributions with skew and kurtosis values all < |1.0|.

Growth Model

Fit statistics for growth models are shown in Appendix S5. The best fit to the data was provided by a LGM-SR. This model used linear through quartic growth factors, indicating that the sample average rate of change in depression was different at different ages. It included significant variances for the linear, quadratic, and cubic growth factors (the quartic term was constrained to be equal across children), indicating there was variation in children's linear and curvilinear rates of change in depression. The residuals were moderately positively correlated (standardized autoregression parameter [AR] = 0.30), meaning that the average child had a small to moderate amount of renitent responding. Despite significant variation in trajectories across participants, the model implied mean did not change much from baseline scores, with the sample average peak and trough .23 *SD* apart.

Mixture Model

Fit statistics for mixture models are presented in Appendix S6. Information criteria were unhelpful in model selection because they all decreased without signs of leveling off with additional classes. Consequently, the final model was selected based on likelihood ratio tests (Lo-Mendell-Rubin adjusted likelihood ratio test), visual inspection of fit to individual level data, and interpretability. The best performing model based on these criteria was a six-class model with intercept and slope variances constrained equal across classes, quadratic and higher degree polynomial variances constrained at zero, a diagonal covariance matrix (i.e., the covariation between intercept and slope was fully explained by latent class membership), and an autoregressive parameter that varied across classes.

Table 1

Means, Standard Deviations, and Correlations for Study Variables

Depression symptom scores by age	n (% nonmissing)	M	SD	Min/max	1	2	3	4	5	6	7	8	9	10	11	12
1. 3.9 years	9,464 (81.3)	.76	.29	0.03/1.88	—	—	—	—	—	—	—	—	—	—	—	—
2. 6.8 years	8,400 (72.2)	.77	.35	0.05/2.03	.79	—	—	—	—	—	—	—	—	—	—	—
3. 8.1 years	7,794 (67.0)	.84	.36	0.07/2.06	.72	.91	—	—	—	—	—	—	—	—	—	—
4. 9.6 years	8,031 (69.0)	.76	.37	0.00/2.14	.64	.84	.88	—	—	—	—	—	—	—	—	—
5. 11.7 years	7,392 (63.5)	.74	.38	0.05/2.24	.56	.76	.79	.84	—	—	—	—	—	—	—	—
6. 13.1 years	7,122 (61.2)	.74	.39	0.04/2.24	.55	.75	.76	.81	.87	—	—	—	—	—	—	—
7. 16.5 years	5,695 (48.9)	.75	.38	0.02/2.33	.53	.67	.71	.71	.74	.79	—	—	—	—	—	—

Social predictor variables	n (% nonmissing)	%	Count	Min/max	1	2	3	4	5	6	7	8	9	10	11	12
8. male	11,641 (100.0)	51.0	5,938	0/1	-.05	-.05	-.07	-.07	-.09	-.14	.01	—	—	—	—	—
9. Maternal psy.	8,431 (72.4)	23.1	1,945	0/1	.29	.31	.29	.28	.27	.26	.00	.01	—	—	—	—
10. Single parent	8,239 (70.8)	12.1	999	0/1	.10	.11	.11	.10	.09	.11	-.03	.00	.25	—	—	—
11. Neighborhood dis.	8,578 (73.7)	13.2	1,135	0/1	.19	.19	.19	.19	.18	.16	.12	-.03	.31	.30	—	—
12. Child abuse	9,019 (77.5)	6.1	551	0/1	.16	.16	.15	.15	.14	.14	-.01	-.08	.16	.07	.15	—
13. Mother's education[a]	10,583 (90.9)	—	—	1/4	-.06	-.06	-.07	-.08	-.07	-.08	-.01	-.08	-.08	-.22	-.16	.09
Less than O-level		26.9	2,846	—	—	—	—	—	—	—	—	—	—	—	—	—
O-level		35.3	3,736	—	—	—	—	—	—	—	—	—	—	—	—	—
A-level		23.8	2,520	—	—	—	—	—	—	—	—	—	—	—	—	—
Degree or above		14.0	1,481	—	—	—	—	—	—	—	—	—	—	—	—	—

Note. psy. = psychopathology; dis. = disadvantage.
[a] Mother's education was a 4-point ordinal scale. In the British educational system, O-level exams represent 11 years of secondary education and determine student readiness for more advanced (A-level) coursework. Depression scores were log-transformed. Participants were included in the count of nonmissing if they had at least one complete parcel of the Strengths and Difficulties Questionnaire or Short Mood and Feelings Questionnaire. Polyserial and polychoric correlations were calculated for continuous and categorical items, respectively, using full information maximum likelihood estimation.
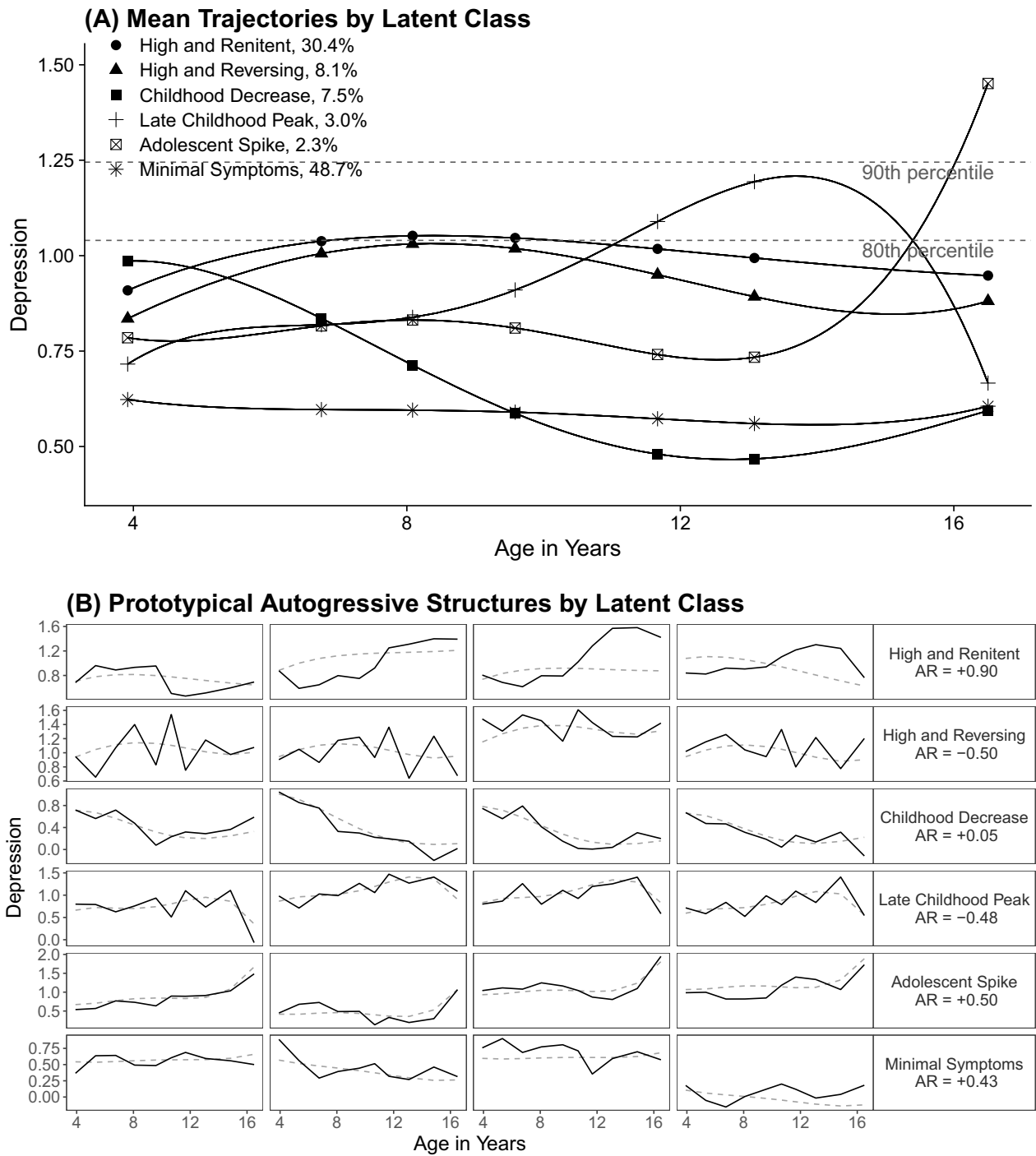
(A) Mean Trajectories by Latent Class

(B) Prototypical Autogressive Structures by Latent Class

Figure 3. (A) represents the average latent trajectory for members of each latent class. (B) represents prototypical patterns of variation around each trajectory (generated with a parametric bootstrap). Gray dashed lines represents person-specific mean trajectories. Solid black lines represent prototypical patterns of variation as captured by the structured residual. AR = autoregressive structured residual.

Latent class trajectories are shown in Figure 3A and prototypical patterns of responding, generated from a parametric bootstrap, are shown in Figure 3B (the associated growth parameters for the 1 and 6 class models are provided in Appendix S7).

To ease interpretation, Figure 3A includes a reference line at the 90th percentile depression score, which is often used as a rough proxy for a clinical diagnosis of depression (Zavos, Rijsdijk, Gregory, & Eley, 2010). Raw trajectories of individuals in each
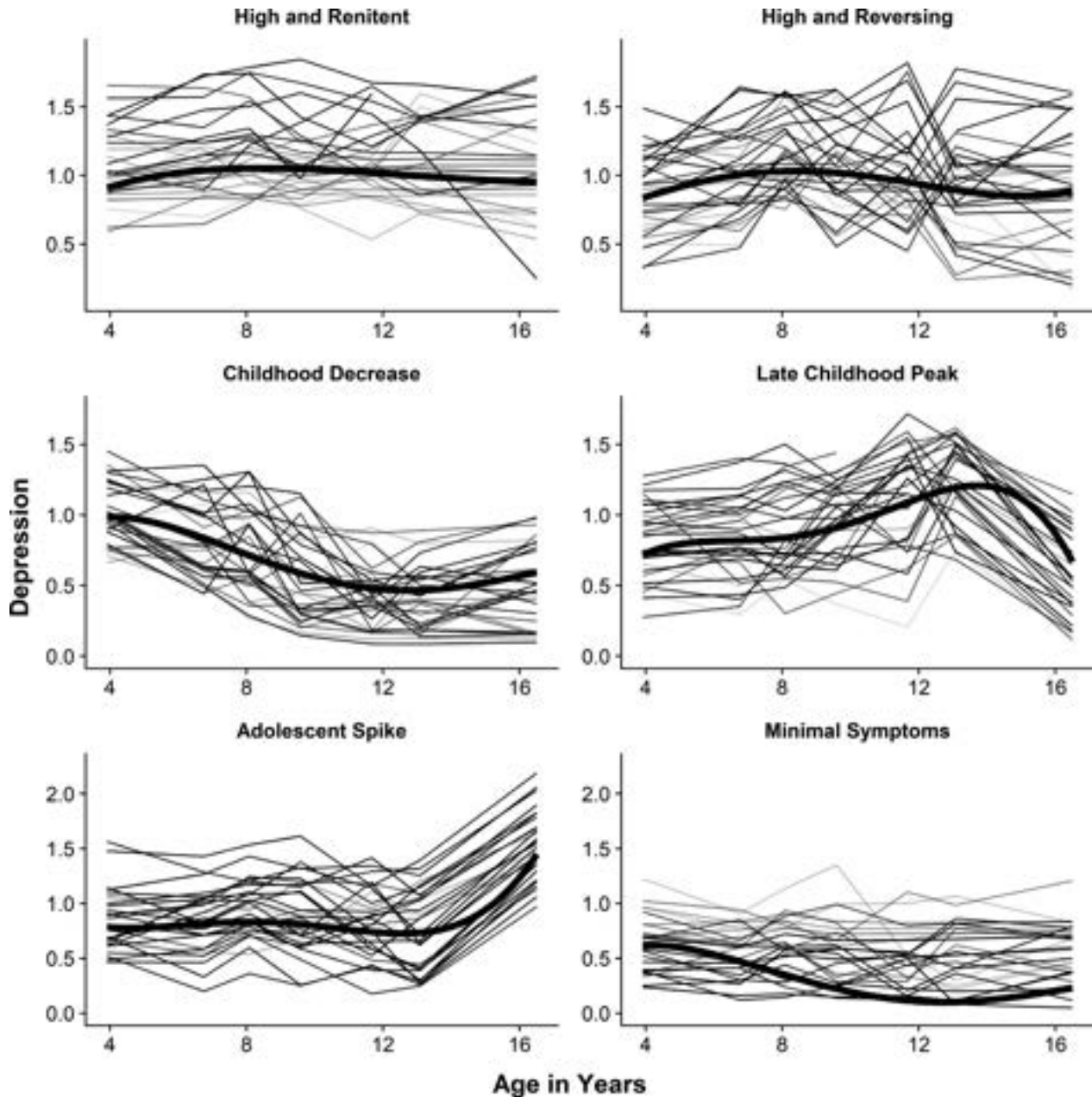
Figure 4. Thirty participants were randomly sampled from each class. Thick lines represent class average trajectories. Line opaqueness represents classification uncertainty, with more opaque lines representing higher classification certainty.

latent class are shown in Figure 4. Classification precision of individuals into latent classes was moderately low (entropy = .67), perhaps because autoregressive terms have considerable error at the individual level with just seven observed time points. However, in samples as large as the present one, autoregressive terms are still accurately estimated at the group level (Schultzberg & Muthén, 2018).

Classes were more clearly separated by their trajectories and patterns of responding than by baseline symptom levels. Developmental trajectories varied considerably across classes. Approximately half of all children experienced stable, low symptoms with moderate levels of renitent responding across childhood and adolescence (*Minimal Symptoms* class, 48.7%, $n = 6{,}543$, AR = 0.43). The next two largest classes experienced moderately high

symptoms and were primarily differentiated by their patterns of responding. The *High and Renitent* class (30.4%, *n* = 3,129) exhibited the strongest levels of renitent responding in the sample (AR = 0.90), meaning that if a life event caused them to move above or below their typical depression level, they tended to stay there for very long periods of time. In contrast to the *High and Renitent* class, the *High and Reversing* class (8.1%, *n* = 618) had nearly the same mean trajectory, but experienced high levels of reversing responding (AR = −0.50), as they experienced sharp oscillations around their estimated mean levels. The *Childhood Decrease* class (7.5%, *n* = 756) experienced high symptoms at an early age, but these symptoms came down gradually over the course of early and middle childhood and remained at a low level over adolescence; their symptom changes had minimal carry-over across waves (AR = 0.05). The final two classes were separated by their trajectories over middle childhood and adolescence. The *Late Childhood Peak* class (3.0%, *n* = 334) experienced an increase in symptoms, primarily through middle childhood, then a large decrease in symptoms over adolescence; this class was also characterized by reversing responding. The *Adolescent Spike* class (2.3%, *n* = 261) remained relatively stable at a low-to-moderate level of symptoms until adolescence, at which point their symptoms spiked to the highest average levels observed in the study.

Social Predictors of Class Membership

Social predictors of class membership are presented in Table 2, and a bar chart of corresponding probabilities of class membership across predictors is shown in Appendix S8. Likelihood ratio tests indicated that each variable was a statistically significant predictor of class membership. Generally, maternal psychopathology, neighborhood disadvantage, childhood abuse, and female sex were associated with an increased risk of belonging to a group other than the minimal symptoms group. Maternal psychopathology conveyed the highest risk, with the probability of being in the *Minimal Symptoms* class dropping by more than half from 57% for those with no maternal psychopathology to 24% for those with maternal psychopathology. The pattern of findings for maternal education was somewhat more complex. Higher levels of maternal education were associated with lower odds of being in the *High and Renitent* group as compared to any other group, but—counterintuitively—higher levels of maternal education were also associated with

increased risk of being in any group as compared to the *Minimal Symptoms* group (excluding the *High and Renitent* group).

Two different sets of comparisons were particularly meaningful. The first was between the *High and Renitent* versus *High and Reversing* classes, because those classes both had high symptom levels and were distinguished primarily by differences in response to time-anchored events. The second was between *Late Childhood Peak* versus *Adolescent Spike* classes, because those classes were distinguished primarily by the timing of a spike in depression symptoms. The *High and Renitent* group was differentiated from the *High and Reversing* group only by maternal education, with higher levels of maternal education associated with higher probabilities of being in the *High and Reversing* group. Although the *High and Renitent* group has more statistically significant effects when contrasted against other latent classes, this is likely driven by power, not necessarily a substantive difference between the two classes. The *High and Renitent* group has more than three times the sample size as the *High and Reversing* group, so has the power to detect smaller effect sizes that would be nonsignificant for the *High and Renitent* group. The *Late Childhood Peak* and *Adolescent Spike* classes were differentiated only by sex; males had three times the odds of females of spiking in adolescence as compared to late childhood.

Sensitivity Analysis

To understand how using a GMM-SR changed the characterization of developmental heterogeneity compared to using more traditional GMMs, the final model was compared to a traditional six-class GMM with the structured residuals removed. Similar developmental trajectories were observed across both parameterizations, but the traditional GMM had classes characterized by more extreme scores. Notably, in the traditional GMM, the group with consistently high symptom levels plateaued at more extreme scores but was smaller in size (11% vs. 30%) than the analogous group in the GMM-SR. This result suggests that the traditional GMM separated a group of children with homogeneous developmental trajectories into more extreme high and low groups because it ignored the potential for patterns of responding to cause depression scores to carry over across waves (similar to Figures 2C–2D). A side-by-side comparison of the two models is available in Appendix S9.

Table 2
Predictors of Latent Class Membership

Predictor	Referent class = minimal symptoms		Referent class = adolescent spike		Referent class = late childhood peak		Referent class = childhood decrease		Referent class = high and reversing	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
High and renitent										
Male	**0.79**	**[0.69, 0.91]**	**2.57**	**[1.84, 3.58]**	0.81	[0.61, 1.06]	**0.64**	**[0.52, 0.79]**	1.11	[0.87, 1.40]
Maternal psychopathology	**5.12**	**[4.30, 6.11]**	**1.63**	**[1.18, 2.26]**	**2.22**	**[1.61, 3.05]**	**2.02**	**[1.60, 2.54]**	1.19	[0.94, 1.51]
Single parent	**1.35**	**[1.09, 1.66]**	1.52	[0.95, 2.42]	**1.86**	**[1.15, 3.02]**	1.29	[0.94, 1.78]	1.03	[0.76, 1.42]
Neighborhood disadvantage	**2.11**	**[1.72, 2.60]**	1.67	[1.05, 2.65]	1.25	[0.85, 1.83]	**1.51**	**[1.12, 2.03]**	1.16	[0.86, 1.56]
Abuse	**2.39**	**[1.80, 3.18]**	1.23	[0.68, 2.21]	1.28	[0.76, 2.14]	1.51	[1.01, 2.24]	1.29	[0.82, 2.03]
Mother's Ed: 0-level	**0.80**	**[0.67, 0.96]**	0.71	[0.46, 1.08]	0.72	[0.48, 1.06]	0.77	[0.58, 1.03]	**0.65**	**[0.47, 0.90]**
Mother's Ed: A-level	**0.67**	**[0.55, 0.82]**	**0.38**	**[0.25, 0.58]**	**0.46**	**[0.31, 0.69]**	**0.58**	**[0.43, 0.79]**	**0.45**	**[0.32, 0.64]**
Mother's Ed: degree	**0.75**	**[0.59, 0.95]**	**0.40**	**[0.24, 0.65]**	**0.32**	**[0.21, 0.50]**	**0.42**	**[0.30, 0.59]**	**0.34**	**[0.23, 0.49]**
High and reversing										
Male	**0.71**	**[0.57, 0.89]**	**2.32**	**[1.59, 3.38]**	0.73	[0.53, 1.00]	**0.58**	**[0.44, 0.76]**	—	—
Maternal psychopathology	**4.30**	**[3.37, 5.48]**	1.37	[0.94, 1.98]	**1.86**	**[1.29, 2.68]**	**1.69**	**[1.26, 2.26]**	—	—
Single parent	1.30	[0.96, 1.77]	1.47	[0.87, 2.48]	1.80	[1.06, 3.08]	1.25	[0.84, 1.86]	—	—
Neighborhood disadvantage	**1.82**	**[1.35, 2.46]**	1.44	[0.86, 2.41]	1.08	[0.69, 1.69]	1.30	[0.89, 1.89]	—	—
Abuse	**1.86**	**[1.18, 2.92]**	0.95	[0.47, 1.92]	0.99	[0.52, 1.87]	1.17	[0.68, 2.00]	—	—
Mother's Ed: 0-level	1.24	[0.91, 1.70]	1.09	[0.66, 1.80]	1.10	[0.69, 1.77]	1.19	[0.81, 1.76]	—	—
Mother's Ed: A-level	**1.49**	**[1.07, 2.07]**	0.85	[0.52, 1.40]	1.03	[0.63, 1.66]	1.29	[0.86, 1.94]	—	—
Mother's Ed: degree	**2.24**	**[1.57, 3.21]**	1.18	[0.67, 2.08]	0.96	[0.58, 1.60]	1.27	[0.82, 1.96]	—	—
Childhood decrease										
Male	1.24	[1.01, 1.53]	**4.02**	**[2.77, 5.82]**	1.26	[0.93, 1.72]	—	—	—	—
Maternal psychopathology	**2.54**	**[1.97, 3.29]**	0.81	[0.56, 1.18]	1.10	[0.77, 1.58]	—	—	—	—
Single parent	1.04	[0.75, 1.45]	1.17	[0.68, 2.01]	1.44	[0.84, 2.48]	—	—	—	—
Neighborhood disadvantage	1.40	[1.02, 1.93]	1.11	[0.65, 1.88]	0.83	[0.53, 1.30]	—	—	—	—
Abuse	1.59	[1.04, 2.44]	0.82	[0.42, 1.60]	0.85	[0.47, 1.54]	—	—	—	—
Mother's Ed: 0-level	1.04	[0.78, 1.39]	0.91	[0.56, 1.48]	0.93	[0.59, 1.45]	—	—	—	—
Mother's Ed: A-level	1.15	[0.85, 1.56]	0.66	[0.41, 1.07]	0.79	[0.50, 1.25]	—	—	—	—
Mother's Ed: degree	**1.77**	**[1.28, 2.45]**	0.93	[0.54, 1.61]	0.76	[0.47, 1.23]	—	—	—	—
Late childhood peak										
Male	0.98	[0.76, 1.27]	**3.18**	**[2.13, 4.76]**	—	—	—	—	—	—
Maternal psychopathology	**2.31**	**[1.67, 3.21]**	0.74	[0.48, 1.13]	—	—	—	—	—	—
Single parent	0.72	[0.45, 1.16]	0.81	[0.43, 1.55]	—	—	—	—	—	—
Neighborhood disadvantage	**1.69**	**[1.15, 2.48]**	1.33	[0.75, 2.37]	—	—	—	—	—	—
Abuse	**1.87**	**[1.11, 3.15]**	0.96	[0.46, 2.02]	—	—	—	—	—	—
Mother's Ed: 0-level	1.12	[0.77, 1.64]	0.99	[0.57, 1.71]	—	—	—	—	—	—
Mother's Ed: A-level	1.45	[0.99, 2.13]	0.83	[0.49, 1.43]	—	—	—	—	—	—
Mother's Ed: degree	**2.33**	**[1.55, 3.50]**	1.23	[0.67, 2.24]	—	—	—	—	—	—

Table Table 2
Continued

Predictor	Referent class = minimal symptoms		Referent class = adolescent spike		Referent class = late childhood peak		Referent class = childhood decrease		Referent class = high and reversing	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
Adolescent spike										
Male	**0.31**	**[0.22, 0.43]**	—	—	—	—	—	—	—	—
Maternal psychopathology	**3.15**	**[2.25, 4.40]**	—	—	—	—	—	—	—	—
Single parent	0.89	[0.55, 1.43]	—	—	—	—	—	—	—	—
Neighborhood disadvantage	1.27	[0.79, 2.03]	—	—	—	—	—	—	—	—
Abuse	1.95	[1.07, 3.56]	—	—	—	—	—	—	—	—
Mother's Ed: 0-level	1.14	[0.75, 1.74]	—	—	—	—	—	—	—	—
Mother's Ed: A-level	**1.74**	**[1.16, 2.63]**	—	—	—	—	—	—	—	—
Mother's Ed: degree	**1.90**	**[1.17, 3.08]**	—	—	—	—	—	—	—	—

Note. "Less than O-level" was the reference group. Block tests indicated that each variable was a statistically significant predictor of class membership; Single Parent Household was significant at $p = .040$; all other variables were significant at $p < .0001$. Bolded estimates were statistically significant after using a Benjamini–Hochberg correction with the false discovery rate set to .05. The reference group for mother's education was < 0-level. Ed = education.

Discussion

This study characterized subgroups of children by their differing developmental trajectories of depression—and differing patterns of responding to life events—from age 4 through 16.5, in a large, prospective, population-based sample. To our knowledge, this was the first GMM study to disentangle age-dependent and age-independent patterns of symptom change using class-specific structured residuals. This study also included a 13-year follow-up period, a wider age span than nearly all previous GMM studies of depression. Overall, the primary findings included identification of potentially sensitive periods for developing depression in both early childhood and adolescence as well as the importance of patterns of responding to life events for explaining the highest levels of childhood depression. Results also suggested robust but complex relationships between social factors of depression and latent class.

One striking difference from previous research is that the classes in this study with consistently high symptom levels were larger in size, but had lower average symptom levels than observed elsewhere, indicating that those at risk for severe depression were drawn from a relatively wide pool of children who experienced varying periods of higher and lower symptoms rather than a smaller pool of children with consistently high symptoms. With a few exceptions, average depression scores in even the most extreme latent class were well below the 90th percentile, which is typically used as a rough proxy for a clinical diagnosis of depression (Zavos et al., 2010). A sensitivity analysis using traditional GMM methods, which excluded the autoregressive parameter that captured renitent and reversing responding (Appendix S9), found latent classes similar to those identified in previous ALSPAC work (Edwards et al., 2014; Rice et al., 2019), with 75% of children in a minimal symptoms group, and those in the higher symptoms groups showing more extreme scores than the high symptoms groups observed in the main analyses. Thus, explicitly modeling renitent and reversing responding leads to the conclusion that the highest levels of depression were better explained by transient states than persistent traits.

Two extremes of responding were evident in these data. The *High and Renitent* and *High and Reversing* groups had nearly identical mean trajectories but were clearly separated by their autoregressive parameters. For the *High and Renitent* group, experiencing a perturbation above their typical depression

level at any given observation occasion was strongly predictive of remaining above their typical level at the next observation. Consequently, these children spent longer continuous periods of time either above or below their expected averages. The *High and Reversing* group had more frequent symptom oscillations, meaning that if these children were below their typical level at one time point, they were likely to be above it at the next time point and vice versa. Together, these two classes can be thought of as comprising children with dynamic risk for depression, with the *High and Renitent* group more likely to experience longer episodes, and the *High and Reversing* group more likely to experience frequent episodes. This risk is termed dynamic because it occurred independently of child age, and was instead driven by unmodeled, trait-level processes that could include both biological factors and characteristic ways of coping with life events. Although the strongest extremes of responding were evident in the classes with the highest average levels of depression, neither pattern of responding should necessarily be considered maladaptive. In fact, both types of responding could signal adaptive or maladaptive responses at different times because any individual is as likely to have a positive residual as a negative residual at any given time. For example, when a person deviates above their expected trajectory, renitent responding is probably harmful because it means they are more likely to continue to experience higher symptoms at the next time point, whereas when a person deviates below their expected trajectory, renitent responding is probably helpful because it means they are more likely to continue to experience lower symptoms at the next time point. The relative merits of response patterns remain unknown, and future research may help to clarify.

This study builds on the literature showing that certain periods of development have increased importance in shaping the course of depressive symptoms. In contrast to the dynamic risk groups described earlier, there appear to be two inflection points of *age-specific* risk for a minority of the sample, one occurring around age 10 and peaking at age 13 (*Late Childhood Peak* group; 3.0% of the sample), and another beginning after age 13 and peaking at age 16.5 (*Adolescent Spike* group; 2.3% of the sample). It is notable that these inflection points correspond with a heightened period of internal transitions (i.e., puberty) and external transitions (i.e., social environmental changes, such as high stakes academic testing and more complex friendships and peer expectations) that previous studies

have linked to the onset of depression (Graber et al., 1997). Sex differences in the present sample correspond closely to different pubertal timing between girls and boys. Girls, who begin to experience puberty as early as age 10, were much more likely to be in the *Late Childhood Peak* group that experienced increasing symptoms between ages 10–13. Boys, who begin to experience puberty several years later—as early as age 12—were much more likely to belong to the *Adolescent Spike* group that experienced increases between ages 13–16. These results are consistent with those of Kwong et al. (2019) who, analyzing ALSPAC data with a different methodology, found that ages 13 and 16 had the peak velocity of changes in depression symptoms for boys and girls, respectively (Kwong et al., 2019). The current results suggest the changes that led to these peak velocities were set in motion 3–5 years prior to the peak velocities being achieved and may have been driven by a small subset of the population. Moreover, the combined size of these two age-specific risk groups was just 5.3% of the sample, whereas 38.5% experienced dynamic risk, suggesting that most of the risk for depression was driven by dynamic responding to environmental events.

In contrast to the sharp inflection points described earlier, the *Childhood Decrease* group had high levels of symptoms early in childhood that decreased over development and into adolescence. Compared to the *High and Renitent* group, children in this class were more likely to be male, have more educated mothers, and less likely to have maternal psychopathology. The contrast to the *High and Renitent* group may reflect early risk that was buffered by protective factors of higher maternal education and lower history of maternal psychopathology.

Three of the six social factors examined in this study—maternal psychopathology, neighborhood disadvantage, and child abuse—were strongly related to differences between the lowest symptoms class (*Minimal Symptoms*) and the highest symptom classes (*High and Renitent*, *High and Reversing*), with odds ratios ranging from 2 to 5. Coming from a single-parent household generated a similar pattern of effects, but with a much smaller effect size. This pattern of findings is consistent with the stress sensitization hypothesis, that exposure to early childhood adversity increases the probability of developing depression in response to any given stressor (Hammen, Henry, & Daley, 2000).

The role of maternal education in determining developmental trajectories of depression appears complex. On one hand, children whose mothers

had higher levels of education were less likely to belong to the *High and Renitent* group than any other group, indicating a protective effect consistent with previous research (Allen, Balfour, Bell, & Marmot, 2014). On the other hand, children with higher maternal education also had lower odds of belonging to the *Minimal Symptoms* group than any other group (except for the *High and Renitent* group), indicating that although they were protected from the *High and Renitent* group, they faced higher risks of experiencing depression in each of the sensitive periods identified by this study.

One possibility for explaining the protective effects of education—that is, the movement from the *High and Renitent* class to other classes with less stability in depression level—is that education level, an indicator of social class, captures the level of access to and quality of resources families can use to buffer their children's risk of depressive symptoms. Notably, maternal education was the only social determinant to differentiate the *High and Renitent* class from the *High and Reversing* class. Although it is important to avoid overinterpreting this finding because maternal education was generally predictive of belonging to any group other than the *High and Renitent* group, one possibility is that low socioeconomic status predicts poorer child socioemotional development and adaptive functioning (Bradley & Corwyn, 2002), and these poor socioemotional skills lead to maladaptive coping behaviors that increase children's vulnerability to depression in the face of stressful life events and other risk factors.

Why children with higher maternal education were less likely to be in the *Minimal Symptoms* class than any other class is puzzling. One possibility is that given the relative lack of social mobility and perceived importance of education for class mobility (Goldthorpe, 2016), children of more educated parents feel more pressure to achieve academically from a young age, leading to higher levels of stress. Alternatively, parents with higher levels of economic disadvantage may be facing myriad other stressors and have less opportunity to identify smaller shifts in their children's emotions and behavior as they emerge over time.

Limitations

This study was limited in terms of both the outcome measurement (examining only maternal report of depression symptoms) and range of etiologic predictors examined. Although we used latent variable models to attempt to combine information

across mother and child reports of symptoms, the only wave where reporter data overlapped—age 16.5—had low correlations between parent and child measures ($r = .36$), suggesting these instruments measured different constructs. It is possible that the strong relationship between maternal psychopathology and childhood depression risk may be explained by a measurement artifact whereby depressed mothers were more likely to rate their children as depressed. The inclusion of a wider range of etiologic predictors in future research could enable the detection of more meaningful distinctions between classes. For example, a combining of behavioral measures, such as parent coping styles, with biological measures, such as child polygenic risk scores, could shed important light on the causes of differential patterns of responding.

A second limitation of this study is the length of time between observations. The theory of renitent and reversing responding draws from clinical observation of recurrent major depression. More frequent measurement occasions would be ideal for understanding the duration of depression episodes with more precision. Nonetheless, the present findings illustrate that patterns of responding to life events that transact over longer periods of time have important implications for how depression functions across childhood and adolescence.

A third limitation of this study relates to congeniality between the imputation model and measurement model. Data were imputed from a homogeneous overall population measurement model but analyzed in a mixture model. Imputing data under a more restrictive model is a conservative choice and would tend to attenuate sample-based heterogeneity with regards to evidence of mixtures in the population. We examined missing data patterns across classes and almost certainly observed attenuation of the estimated size of the *Adolescent Spike* class, which contained no individuals with missing data at the final wave where symptom levels spiked. We also observed that the *High and Renitent* class had more missingness (42%) as compared to the *High and Reversing* class (26%). To explore the possibility that a lack of sensitivity to the autocorrelation in the imputation model may have been driving the separation of mixture components by the autocorrelations, we compared the autocorrelation structures in the full, imputed data set to pairwise-present data. There were trivial differences, indicating that heterogeneity with respect to renitent and reversing responding was not an artifact of the missing data pattern or the missing data imputation.

Finally, this sample is limited in terms of geographic diversity. Given high levels of study enrollment from families in the Bristol area, it is a strongly representative sample; however, the extent to which the sensitive period findings will generalize outside of the United Kingdom, with its particular social class structure and socially specific developmental milestones, remains unclear. Cross-cultural replication may be helpful for disentangling the effects of age-related sensitive periods from those related to socially imposed developmental milestones, such as educational testing.

Conclusions

This study captured the widest age range to-date in a GMM study of childhood and adolescent depression and was the first GMM-SR study to simultaneously explore systematic age-dependent and age-independent patterns of changes in depression. It identified sensitive periods of age-specific risk at ages 8 and 13 when depression symptoms began to grow for a minority of the sample, whose symptoms would peak 3–5 years later. Most importantly, the risk for experiencing the highest levels of depression was neither age-specific nor drawn from a small pool of persistently depressed children, but instead drawn from a larger pool of children who moved dynamically between higher and lower symptom levels depending on how they responded to life events. From a prevention perspective, these findings suggest that early interventions targeting broad coping processes rather than indicated interventions targeting age-specific developmental events may be most effective for decreasing the burden of depression across the life course.

References

Allen, J., Balfour, R., Bell, R., & Marmot, M. (2014). Social determinants of mental health. *International Review of Psychiatry*, 26, 392–407. https://doi.org/10.3109/09540261.2014.928270

Allen, J. P., Chango, J., Szwedo, D., & Schad, M. (2014). Long-term sequelae of subclinical depressive symptoms in early adolescence. *Development and Psychopathology*, 26, 171–180. https://doi.org/10.1017/s095457941300093x

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

Angold, A., Costello, E. J., Pickles, A., Messer, S. C., Winder, F., & Silva, D. (1995). The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, 5, 237–249.

Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the national comorbidity survey-adolescent supplement: prevalence, correlates, and treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54 (1), 37–44. https://doi.org/10.1016/j.jaac.2014.10.010

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, 83, 871–892. https://doi.org/10.1007/s11336-017-9592-7

Bauer, D. J.(2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757–786. https://doi.org/10.1080/00273170701710338

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3), 338https://doi.org/10.1037/1082-989X.8.3.338

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bolger, N. (1990). Coping as a personality process: a prospective study. *Journal of Personality and Social Psychology*, 59, 525–537. https://doi.org/10.1037/0022-3514.59.3.525

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Cairns, K. E., Yap, M. B., Pilkington, P. D., & Jorm, A. F. (2014). Risk and protective factors for depression that adolescents can modify: A systematic review and meta-analysis of longitudinal studies. *Journal of Affective Disorders*, 169, 61–75. https://doi.org/10.1016/j.jad.2014.08.006

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6, 330–351. https://doi.org/10.1037/1082-989X.6.4.330

Compas, B. E., Orosan, P. G., & Grant, K. E. (1993). Adolescent stress and coping: Implications for psychopathology during adolescence. *Journal of Adolescence*, 16, 331–349. https://doi.org/10.1006/jado.1993.1028

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Curran, P. J., Cole, V. T., Bauer, D. J., Rothenberg, W. A., & Hussong, A. M. (2018). Recovering predictor–criterion relations using covariate-informed factor score estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 860–875. https://doi.org/10.1080/10705511.2018.1473773

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent growth model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*, 879–894. https://doi.org/10.1037/a0035297

Dekker, M. C., Ferdinand, R. F., van Lang, N. D., Bongers, I. L., van der Ende, J., & Verhulst, F. C. (2007). Developmental trajectories of depressive symptoms from early childhood to late adolescence: Gender differences and adult outcome. *Journal of Child Psychology and Psychiatry*, *48*, 657–666. https://doi.org/10.1111/j.1469-7610.2007.01742.x

Dunn, E. C., Soare, T. W., Raffeld, M. R., Busso, D. S., Crawford, K. M., Davis, K. A., . . . Susser, E. S. (2018). What life course theoretical models best explain the relationship between exposure to childhood adversity and psychopathology symptoms: Recency, accumulation, or sensitive periods? *Psychological Medicine*, *48*, 2562–2572. https://doi.org/10.1017/S0033291718000181

Edwards, A. C., Latendresse, S. J., Heron, J., Cho, S. B., Hickman, M., Lewis, G., . . . Kendler, K. S. (2014). Childhood internalizing symptoms are negatively associated with early adolescent alcohol use. *Alcoholism: Clinical and Experimental Research*, *38*, 1680–1688. https://doi.org/10.1111/acer.12402

Ellis, R. E. R., Seal, M. L., Simmons, J. G., Whittle, S., Schwartz, O. S., Byrne, M. L., & Allen, N. B. (2017). Longitudinal trajectories of depression symptoms in adolescence: Psychosocial risk factors and outcomes. *Child Psychiatry and Human Development*, *48*, 554–571. https://doi.org/10.1007/s10578-016-0682-z

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V. J., . . . Marks, J. S. (1998). Relationships of childhood abuse and household dysfunction to many of the leading causes of death in adults: The adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine*, *14*, 245–258.

Franklin, J. C., Jamieson, J. P., Glenn, C. R., & Nock, M. K. (2015). How developmental psychopathology theory and research can inform the research domain criteria (RDoC) project. *Journal of Clinical Child & Adolescent Psychology*, *44*, 280–290. https://doi.org/10.1080/15374416.2013.873981

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., . . . Lawlor, D. A. (2012). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, *42*, 97–110. https://doi.org/10.1093/ije/dys066

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 626,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, *8*, 651–662.

Garvey, M. J., & Schaffer, C. B. (1994). Are some symptoms of depression age dependent? *Journal of Affective Disorders*, *32*, 247–251. https://doi.org/10.1016/0165-0327(94)90088-4

Goldthorpe, J. H. (2016). Social class mobility in modern Britain: Changing structure, constant process. *Journal of the British Academy*, *4*, 89–111. https://doi.org/10.5871/jba/004.089

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Gould, M. S., Greenberg, T., Velting, D. M., & Shaffer, D. (2003). Youth suicide risk and preventive interventions: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 386–405. https://doi.org/10.1097/01.Chi.0000046821.95464.Cf

Graber, J. A., Lewinsohn, P. M., Seeley, J. R., & Brooks-Gunn, J. (1997). Is psychopathology associated with the timing of pubertal development? *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(12), 1768–1776. https://doi.org/10.1097/00004583-199712000-00026

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323. https://doi.org/10.1037/1082-989x.11.4.323

Hammen, C., Henry, R., & Daley, S. E. (2000). Depression and sensitization to stressors among young women as a function of childhood adversity. *Journal of Consulting and Clinical Psychology*, *68*, 782–787. https://doi.org/10.1037/0022-006X.68.5.782

Hegeman, J. M., Kok, R. M., van der Mast, R. C., & Giltay, E. J. (2012). Phenomenology of depression in older compared with younger adults: Meta-analysis. *British Journal of Psychiatry*, *200*, 275–281. https://doi.org/10.1192/bjp.bp.111.095950

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, *15*, 222. https://doi.org/10.1186/s12888-015-0596-5

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901–930. https://doi.org/10.1037/a0038822

Hyde, J. S., Mezulis, A. H., & Abramson, L. Y. (2008). The ABCs of depression: Integrating affective, biological, and cognitive models to explain the emergence of gender differences in depression. *Psychological Review*, *115*, 291–313.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., . . . Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic

analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392, 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126, 454–477. https://doi.org/10.1037/abn0000258

Kuppens, P., Sheeber, L. B., Yap, M. B., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, 12, 283–289. https://doi.org/10.1037/a0025046

Kwong, A. S. F., Manley, D., Timpson, N. J., Pearson, R. M., Heron, J., Sallis, H., . . . Leckie, G. (2019). Identifying critical points of trajectories of depressive symptoms from childhood to young adulthood. *Journal of Youth and Adolescence*, 48, 815–827. https://doi.org/10.1007/s10964-018-0976-5

Lai, H. M., Cleary, M., Sitharthan, T., & Hunt, G. E. (2015). Prevalence of comorbid substance use, anxiety and mood disorders in epidemiological surveys, 1990–2014: A systematic review and meta-analysis. *Drug and Alcohol Dependence*, 154, 1–13. https://doi.org/10.1016/j.drugalcdep.2015.05.031

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: Age at onset, episode duration, and time to recurrence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 33, 809–818. https://doi.org/10.1097/00004583-199407000-00006

Lewinsohn, P. M., Rohde, P., Seeley, J. R., Klein, D. N., & Gotlib, I. H. (2000). Natural course of adolescent major depressive disorder in a community sample: Predictors of recurrence in young adults. *American Journal of Psychiatry*, 157, 1584–1591. https://doi.org/10.1007/s10964-018-0976-5

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Lorenzo-Luaces, L. (2015). Heterogeneity in the prognosis of major depression: From the common cold to a highly debilitating and recurrent illness. *Epidemiol Psychiatr Sci*, 24, 466–472. https://doi.org/10.1017/s2045796015000542

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (pp. 551–611). Oxford, UK: Oxford University Press.

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ)-further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child & Adolescent Psychiatry*, 12(1), 1–8. https://doi.org/10.1007/s00787-003-0298-2

Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders*, 192, 199–211. https://doi.org/10.1016/j.jad.2015.12.030

Muthen, B., & Muthen, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research*, 24, 882–891. https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Author.

Patton, G. C., Olsson, C., Bond, L., Toumbourou, J. W., Carlin, J. B., Hemphill, S. A., & Catalano, R. F. (2008). Predicting female depression across puberty: A two-nation longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 1424–1432. https://doi.org/10.1097/CHI.0b013e3181886ebe

Petterson, S. M., & Albers, A. B. (2001). Effects of poverty and maternal depression on early child development. *Child Development*, 72, 1794–1813. https://doi.org/10.1111/1467-8624.00379

Rice, F., Riglin, L., Thapar, A. K., Heron, J., Anney, R., O'Donovan, M. C., & Thapar, A. (2019). Characterizing developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression. *JAMA Psychiatry*, 76, 306–313. https://doi.org/10.1001/jamapsychiatry.2018.3338

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 495–515. https://doi.org/10.1080/10705511.2017.1392862

Shore, L., Toumbourou, J. W., Lewis, A. J., & Kremer, P. (2018). Review: Longitudinal trajectories of child and adolescent depressive symptoms and their predictors—A systematic review and meta-analysis. *Child and Adolescent Mental Health*, 23, 107–120. https://doi.org/10.1111/camh.12220

Ten Have, M., de Graaf, R., van Dorsselaer, S., Tuithof, M., Kleinjan, M., & Penninx, B. (2018). Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. *Acta Psychiatrica Scand.*, 137, 503–515. https://doi.org/10.1111/acps.12874

Turner, N., Joinson, C., Peters, T. J., Wiles, N., & Lewis, G. (2014). Validity of the Short Mood and Feelings Questionnaire in late adolescence. *Psychological Assessment*, 26, 752–762. https://doi.org/10.1037/a0036572

van Loo, H. M., de Jonge, P., Romeijn, J. W., Kessler, R. C., & Schoevers, R. A. (2012). Data-driven subtypes of

major depressive disorder: A systematic review. *BMC Medicine*, *10*, 156.

Whalen, D. J., Luby, J. L., Tilman, R., Mike, A., Barch, D., & Belden, A. C. (2016). Latent class profiles of depressive symptoms from early to middle childhood: predictors, outcomes, and gender effects. *Journal of Child Psychology and Psychiatry*, *57*, 794–804. https://doi.org/10.1111/jcpp.12518

Whalen, D. J., Sylvester, C. M., & Luby, J. L. (2017). Depression and anxiety in preschoolers: A review of the past 7 years. *Child and Adolescent Psychiatric Clinics of North America*, *26*, 503–522. https://doi.org/10.1016/j.chc.2017.02.006

Zavos, H. M., Rijsdijk, F. V., Gregory, A. M., & Eley, T. C. (2010). Genetic influences on the cognitive biases associated with anxiety and depression symptoms in adolescents. *Journal of Affective Disorders*, *124*(1–2), 45–53. https://doi.org/10.1016/j.jad.2009.10.030

Zisook, S., Lesser, I., Stewart, J. W., Wisniewski, S. R., Balasubramani, G. K., Fava, M., . . . Rush, A. J. (2007). Effect of age at onset on the course of major depressive disorder. *American Journal of Psychiatry*, *164*, 1539–1546. https://doi.org/10.1176/appi.ajp.2007.061017

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Social Determinants Measures

Appendix S2. Measurement Model Invariance Testing

Appendix S3. Longitudinal Design

Appendix S4. Missing Data

Appendix S5. Growth Model Fit Statistics

Appendix S6. Mixture Model Fit Statistics

Appendix S7. Growth Parameters for 1-Class And 6-Class Models

Appendix S8. Bar Chart of Latent Class Membership Across 6 Predictor Variables

Appendix S9. Comparison of 6-class Solutions for GMM versus GMM-SR

SUPPLEMENTAL MATERIALS

Contents

Appendix A

Social factors. We examined the role of five social factors as predictors of trajectory class membership: 1) maternal education (as a measure of socioeconomic position), 2) maternal psychopathology, 3) caregiver physical/emotional cruelty, 4) living in a single parent household, and 5) neighborhood disadvantage. These factors were chosen based on previous studies showing they precede the development of psychopathology in other samples (Petterson & Albers, 2001; Felitti et al., 1998). To ensure these factors preceded the measurement of depressive symptoms, all factors were measured at or before the age of four years. All social factors were coded consistent with a prior ALSPAC study (Dunn et al., 2018).

Maternal education level, used to capture socioeconomic position, was assessed during the maternal interview at 32 weeks gestation. At the time of measurement, achieving "Ordinary Level" (O-level) on the UK national exam was the first step in the pathway to a university education. Scores were entered on a multinomial ordinal scale ranging from: 1=less than O-level, 2= O-level, 3= A-level, and 4 = university degree or higher.

Maternal psychopathology was assessed at child ages 8 months, 1.75 years, and 2.75 years using two valid and reliable measures: (1) the Crown-Crisp Experiential Index Score (CCEI), which includes separate subscales for anxiety and depression (Crown & Crisp, 1979); and (2) the Edinburgh Postnatal Depression Scale (EDPS; Cox, Holden, & Sagovsky, 1987). In addition, reports on maternal suicide attempts in the past year were collected via mailed questionnaire at the same time points. Children were coded as exposed if at least one of the following criteria occurred at any time point: (1) mother reported a suicide attempt; (2) mother had a CCEI depression score greater than 9; (3) mother had a CCEI anxiety score greater than 10; or (4) mother had an EDPS Scale score greater than 12.

Caregiver physical and emotional cruelty. Children were coded as exposed if at the 3.5 years timepoint either the mother or the mother's partner responded "yes" to any of the following items via mailed questionnaire: (1) Your partner was physically cruel to your children; (2) You were physically cruel to your children; (3) Your partner was emotionally cruel to your children; or (4) You were emotionally cruel to your children.

Single parent household. Number of adults in the household was recorded via mailed questionnaire, and a binary variable was generated indicating whether the child lived with a single parent during at least one of the following timepoints: 8 months, 2.75 years, or 3.9 years.

Neighborhood disadvantage was measured at 1.75 and 2.75 years via mailed questionnaires asking mothers to indicate the degree to which the following were problems in their neighborhood: (1) noise from other homes; (2) noise from the street; (3) garbage on the street; (4) dog dirt; (5) vandalism; (6) worry about burglary; (7) mugging; and (8) disturbance from youth. Response options to each item were: 2=serious problem, 1=minor problem, 0=not a problem or no opinion. Items were summed, and children with scores of eight or greater during at least one time point were classified as exposed to neighborhood disadvantage.

Appendix B

Measurement Model and Invariance Testing

Item-Level Analyses

Because the SMFQ included 13 items with a greater range of symptoms, full, item-level invariance analyses were conducted. Factors were modeled with a latent response variable specification (Muthen, 1983), where items were ordinal indicators of an underlying continuous variable. This parameterization was chosen because items had relatively few response points (three levels of responses; see Rhemtulla, Brosseau-Liard, & Savalei, 2012). The SMFQ consisted of three response levels, *not true, sometimes,* and *true.* Thresholds refer to the probability, expressed as a logit, that an item would be endorsed after controlling for the level of the latent variable. Scalar invariance signifies that changes in item endorsement over time are driven entirely by the level of the latent variable. If items measuring a latent variable are treated as scalar invariant when it does not hold, scores of the latent variable are distorted by changes in the items unrelated to the latent variable. If scalar noninvariance was found, results from these item-level models would be used to construct parcels with longitudinally similar patterns of noninvariance.

Although the model demonstrated adequate global fit, the full scalar model was significantly worse by both the chi-square difference test, $\chi^2(74) = 1677.55$, $p < .0001$, as well as Cheung & Rensvold's (2002) more relaxed criterion that changes in the CFI should be $\leq .002$. To conduct partial invariance testing, a minimum of 1 threshold per wave had to be constrained equal to the neighboring wave in order to produce a connected graph and identify the partial scalar model. The initial constraints are depicted by black circles in **Figure B1**. Results of

invariance tests are shown in **Table B1** and the constraints for each test are depicted in **Figure B2**.

Parcel #1 consisted of items where scalar invariance held across both levels of responses across all three waves. That is, after controlling for the level of the latent depression variable, the response for a particular item was equally likely across all three waves. Parcel #1 had full scalar invariance across all waves of items and consisted of items 5, 8, 10, and 12 and broadly reflected negative self-evaluation. Parcel #2 consisted of items 1, 3, 4, 6 and 7, which were mostly emotional symptoms (feeling miserable or unhappy, tired, restless, crying). For these items, *after controlling for the level of depression*, the higher threshold response, *true,* functioned similarly across ages, but the lower threshold response, *sometimes,* was more readily endorsed at 198 months than 140 and 157 months, indicating that after controlling for the latent depression variable, mothers reported their children experiencing moderate levels of negative emotions more frequently at 198 months than they did at younger ages. Parcel #3 consisted of items 2, 9, 11, and 13. Scalar invariance held across 140 and 157 months, but average symptom levels, after controlling for the latent depression variable, were higher at 198 months. These items did not appear to have thematic similarity.

Parcels required a minimum of 3 items. Because the SDQ subscales each had only five items, each subscale was used to form 1 parcel. Item-level analyses on subscales indicated acceptable fit for scalar invariance models for the SDQ-emotions and SDQ-peer difficulties.
Construction of the Measurement Model.

The two SDQ subscales and three SMFQ parcels were included together in the same model and subjected to full invariance testing. Latent means were centered at zero and latent variances were fixed at one in the first wave. A maximum likelihood estimator with robust

5

standard errors was used so that inferences on the measurement model would accommodate nonnormality. The Satorra-Bentler adjustment was used to compare nested models (Satorra & Bentler, 2010).

We could only test metric invariance in the three waves where all measures were present (ages 11.7, 13.1 and 16.5). Consistent with the item-level analyses, preliminary models examining just those three waves indicated that a bifactor model provided the best fit to the data. Moreover, given the heterogeneous range of symptoms assessed and desire for one latent construct that unified them, the bifactor model also provided the substantive interpretation that most cleanly mapped to theory. This measurement model consisted of one general factor underlying all the items, referred to as "depression," and method factors for the SMFQ and each SDQ subscale.

Following metric invariance testing, we constrained factor loadings of the earlier SDQ subscales equal to the later waves. Results indicated full scalar invariance for the SDQ-emotions subscale, and partial scalar invariance for the peer difficulties subscale, where the intercepts of the first three waves (3.9 years, 6.8 years, and 8.1 years) were freely estimated. As expected following the item-level analyses, the two SMFQ parcels constructed from noninvariant items were themselves scalar noninvariant. Nested model comparisons for the configural, metric, full scalar, and partial scalar models can be found in the **Table B2**; for path diagram, see **Figure B1**). Despite a significant chi-square statistic ($\chi^2(165) = 657.02$, $p < .0001$), which is common in models with large samples and many parameters, the fit of the final model with partial scalar invariance was excellent (CFI = .999, TLI = .998, RMSEA = .015) and superior to the fit of the model with full scalar invariance ($\chi^2_{\text{diff}}(5) = 970.50$, $p < .0001$), indicating that some symptoms changed over time in ways that were unrelated to the latent depression variable. Examination of

the size and distribution of residuals indicated that local fit was also acceptable. Factor loadings for the final measurement model are shown in **Table B3**, and latent variable correlations from this model are shown in **Table B4**.

Factor score distributions had moderate levels of skew and kurtosis. Because mixture models on variables with such skew and kurtosis may extract latent classes that characterize properties of the distribution rather than true population-wide heterogeneity (Bauer & Curran, 2003), log-transformed factor scores were used. The resulting estimated factor score distributions were acceptable, with an approximate normal shape and skew and kurtosis values all $< |1.0|$.

Table B1. Nested Model Comparisons for Item-Level SMFQ Partial Scalar Invariance Models

Model	df	χ^2	Δc^2	difftest	CFI	TLI	RMSEA
Full Invariance Models							
Configural/Metric invariance	684	3001.17	---	0.0003	0.983	0.982	0.020
Threshold Invariance	734	3986.29	1677.55	< .0001	0.976	0.976	0.023
Partial Scalar Invariance Models							
Baseline=Model 2-Metric Invariance	684	3001.17	---	---	0.983	0.982	0.020
Partial 0	686	3004.75	2.59	0.27	0.983	0.982	0.020
Partial 1	688	3010.76	6.64	0.036	0.983	0.982	0.020
Partial 2	691	3017.01	5.43	0.143	0.983	0.982	0.020
Partial 3	692	3024.30	10.56	0.001	0.983	0.982	0.020
Partial 4	693	3029.76	19.13	0.0001	0.983	0.982	0.020
Partial 5	697	3050.20	35.58	< .0001	0.983	0.982	0.020
Partial 6	698	3053.87	5.56	0.018	0.983	0.982	0.020
Partial 7	700	3056.18	0.27	0.875	0.983	0.982	0.020
Partial 8	702	3059.05	3.47	0.176	0.983	0.982	0.020
Partial 9	704	3062.09	1.74	0.418	0.983	0.982	0.020
Partial 10	708	3078.98	24.33	0.0001	0.983	0.982	0.020
Partial 11	710	3081.97	0.59	0.744	0.983	0.982	0.020
Partial 12	712	3118.91	76.7	< .0001	0.982	0.982	0.020
Partial 13	714	3124.26	6.02	0.049	0.982	0.982	0.020

Note. This table is interpreted in conjunction with Figure B1, which shows the constraints tested by each model. Chi-square differences between successive models do not equal exact chi-square differences because they were calculated using the Satorra-Bentler adjustment.

Table B2. Fit Statistics for Parcel-Level Partial Scalar Invariance Models

Model	df	χ^2	χ^2 p-value	Satorra-Bentler χ^2 diff	Adjusted χ^2 difference p-value	CFI	TLI	RMSEA	BIC
Waves with all three measures (140 months, 157 months, 198 months)									
Configural	60	263.38	< .0001	---	---	.994	.989	.020	-4,611
Metric	68	328.94	< .0001	58.52	< .0001	.992	.988	.021	-4,567
All waves									
Metric	158	596.74	< .0001	---	---	.999	.998	.015	75,441
Partial Scalar	165	657.03	< .0001	71.43	< .0001	.998	.998	.016	75,448
Full Scalar	170	1353.91	< .0001	989.11	< .0001	.997	.995	.024	76,332

Table B3. Factor Loadings and for the Measurement Model

Age in years	Strengths and Difficulties Questionnaire		Short Mood and Feelings Questionnaire		
	Emotional problems	Peer Difficulties	Parcel #1	Parcel #2	Parcel #3
3.9 years	.69 (.02)	.39 (.01)	---	---	---
6.8 years	.73 (.01)	.48 (.01)	---	---	---
8.1 years	.73 (.01)	.48 (.01)	---	---	---
9.6 years	.74 (.01)	.50 (.01)	---	---	---
11.7 years	.77 (.01)	.49 (.01)	.69 (.01)	.61 (.01)	.66 (.01)
13.1 years	.80 (.01)	.49 (.01)	.69 (.01)	.61 (.01)	.65 (.01)
16.5 years	.81 (.01)	.55 (.01)	.72 (.01)	.68 (.01)	.69 (.01)

Note. Loadings represent standardized values. Standard errors are in parentheses.
Parcel #1 consisted of items 5, 8, 10, and 12.
Parcel #2 consisted of items 1, 3, 4, 6, 7.
Parcel #3 consisted of items 2, 9, 11, 13.

Table B4. Latent Variable Correlations

Age in years	3.9 years	6.8 years	8.1 years	9.6 years	11.7 years	13.1 years	16.5 years
3.9 years	1.00						
6.8 years	0.63	1.38					
8.1 years	0.55	0.79	1.67				
9.6 years	0.46	0.70	0.75	1.61			
11.7 years	0.39	0.62	0.65	0.71	1.66		
13.1 years	0.38	0.59	0.60	0.66	0.75	1.77	
16.5 years	0.33	0.47	0.52	0.51	0.57	0.63	2.07

Note. Variances are on the diagonal.

Figure B1. Constraints on Item-Level SMFQ Partial Invariance Models

Item 13
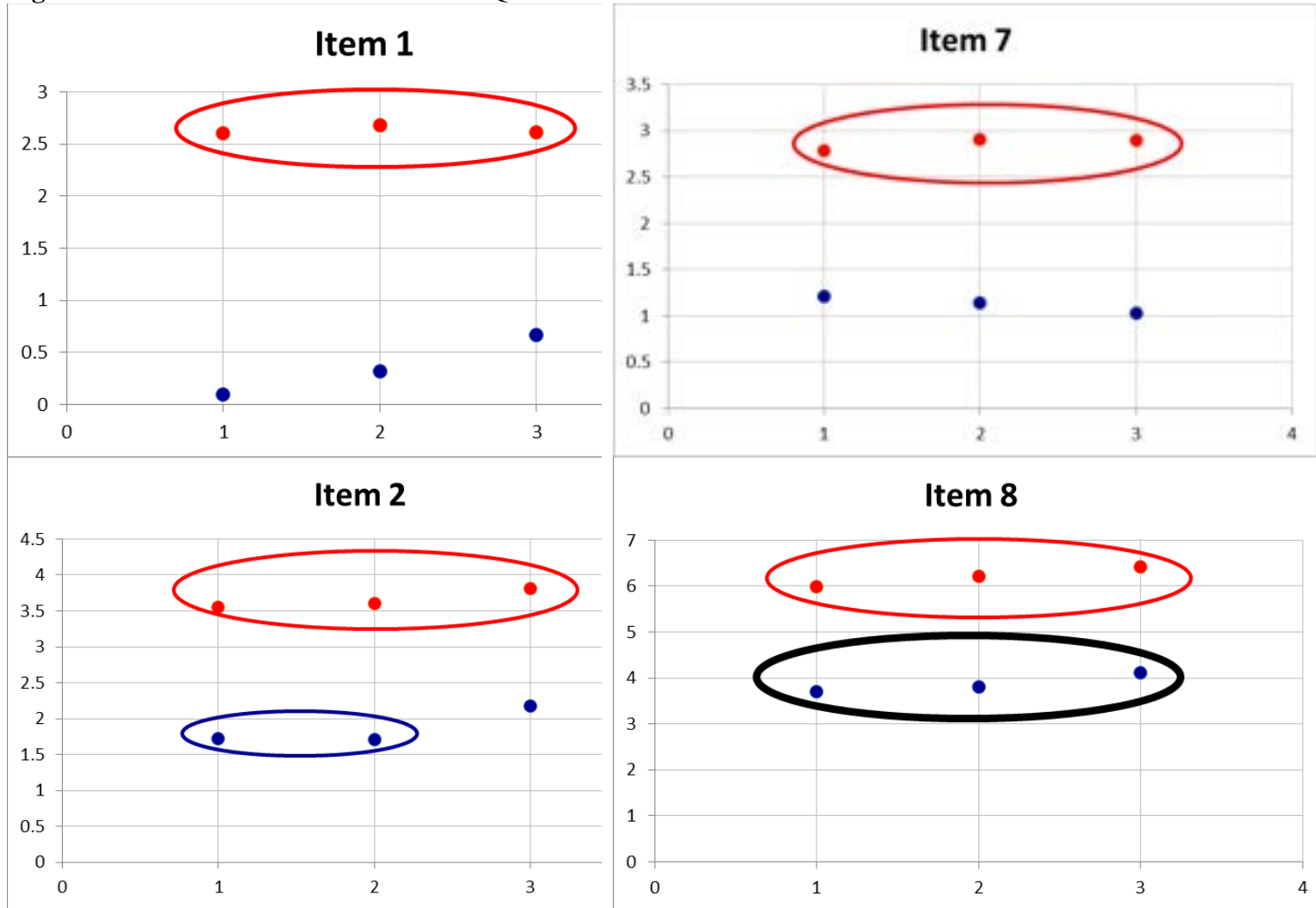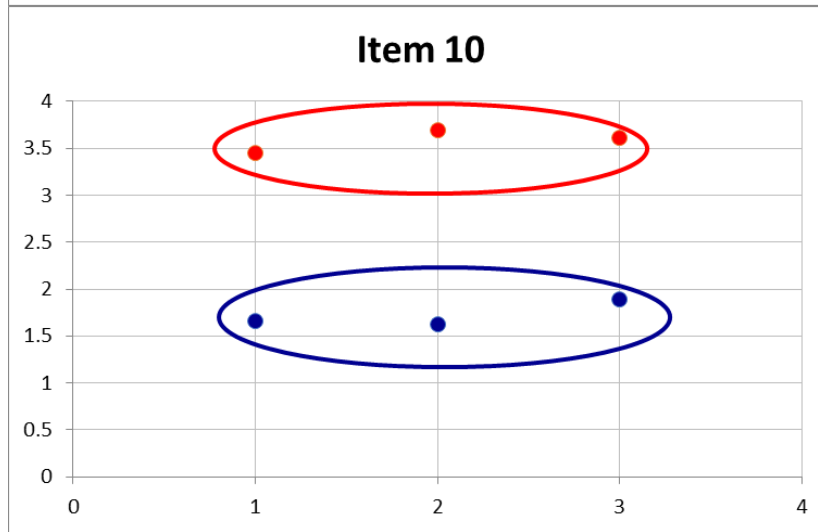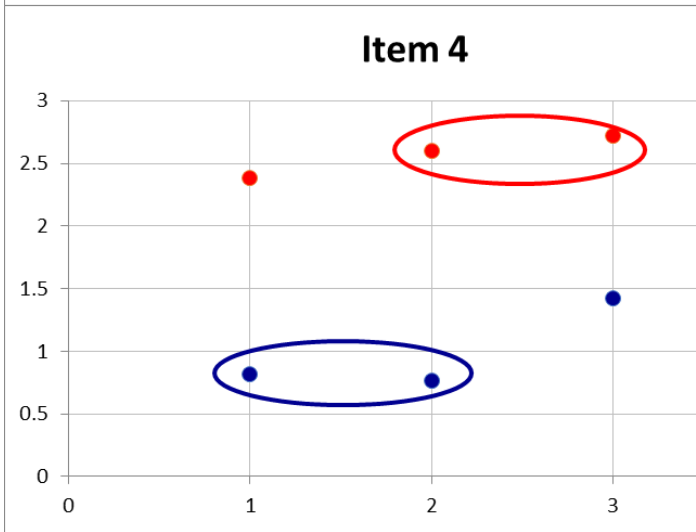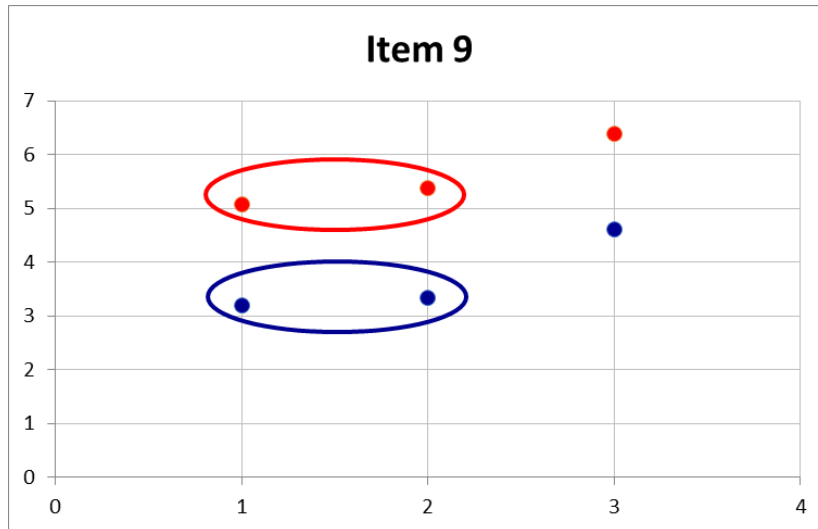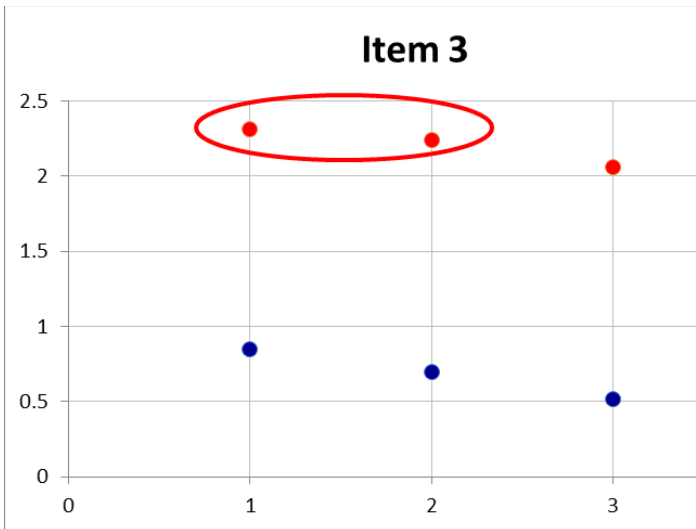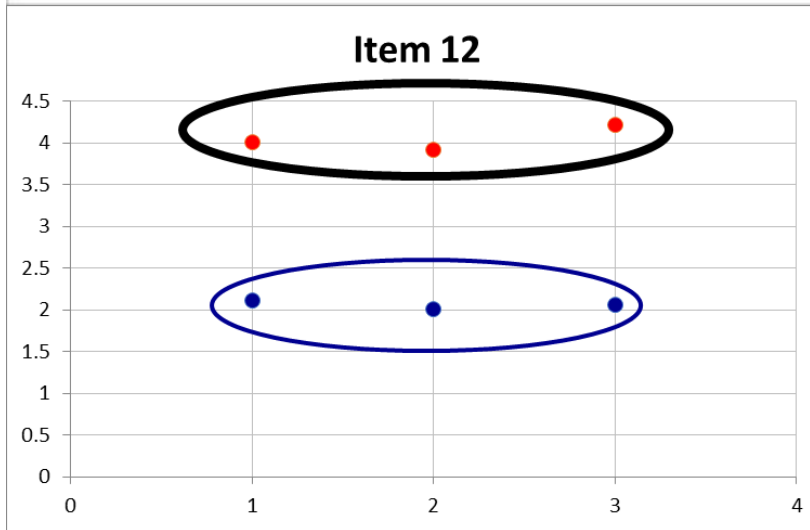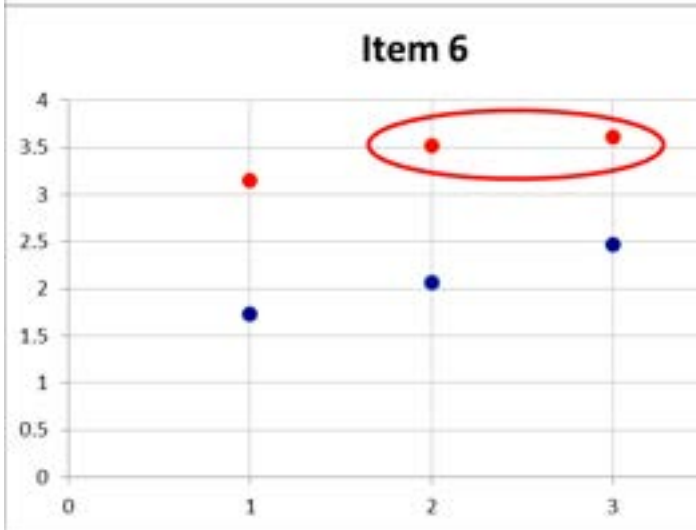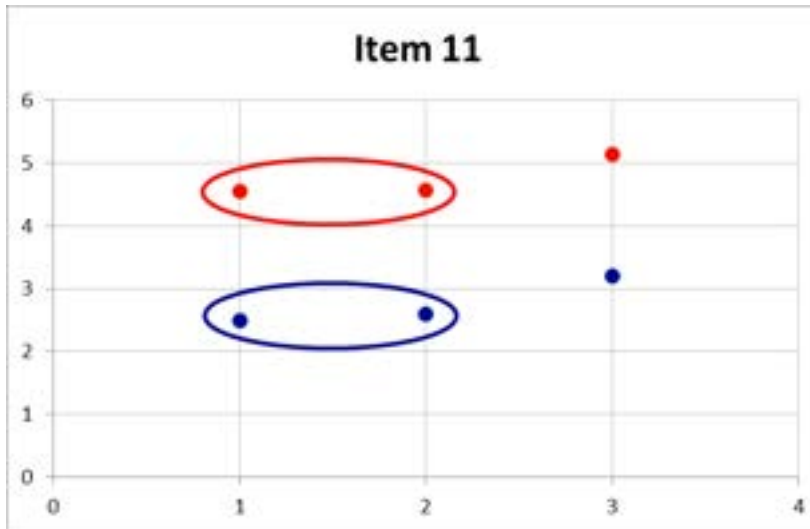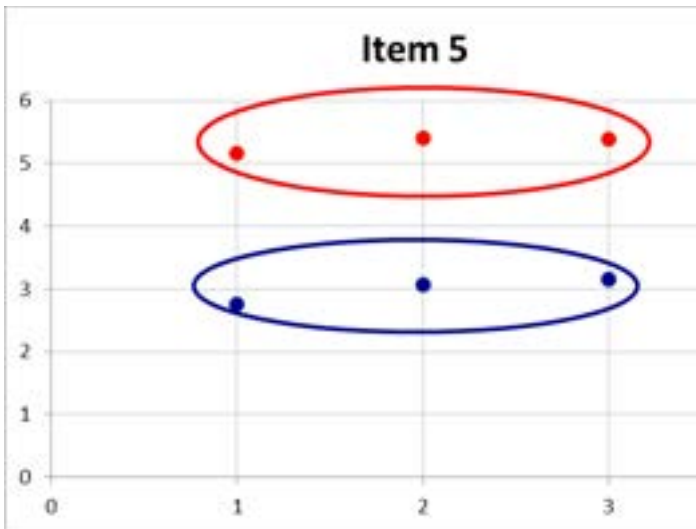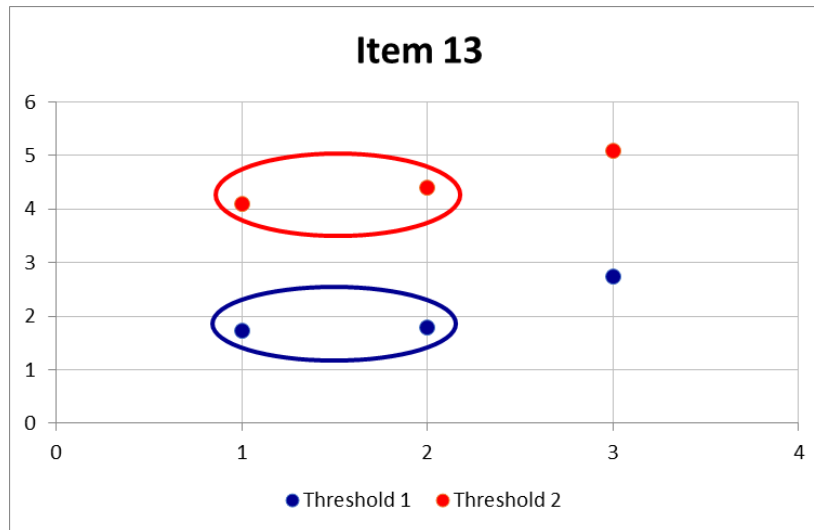
Note. Black circles represent constraints that were used to identify the minimal scalar invariant model (i.e., produce the necessary connected graph for further analysis). Solid circles indicate constraints that were accepted and included in subsequent models. Broken circles represent constraints that were rejected and not imposed for subsequent models.

Figure B2. Path Diagram for Parcel-Level Confirmatory Factor Model
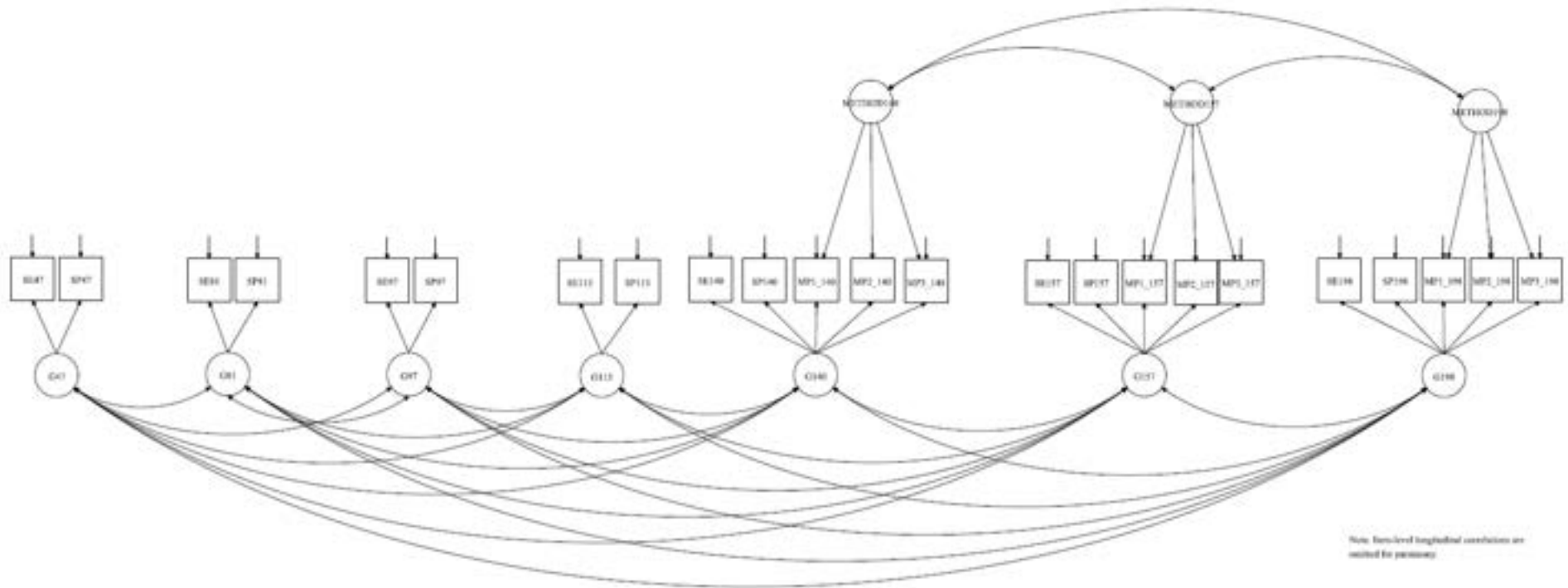
Note. SE = Emotional difficulties subscale of the Strengths and Difficulties Questionnaire. SP = peer difficulties subscale of the Strengths and Difficulties Questionnaire. MP1, MP2, and MP3 represent parcels of the Short Mood and Feelings Questionnaire. Factor loadings are constrained equal for each parcel across waves. Factor variances were scaled to 1 for the first time point and freely estimated at subsequent timepoints.

Table C1. Longitudinal Design Showing Measures and Phantom Variables at Each Wave

Age in years* (months)	Strengths and Difficulties Questionnaire-Emotional Problems	Strengths and Difficulties Questionnaire-Peer Difficulties	Short Mood and Feelings Questionnaire
4 (47)	X	X	
NO MEASURES—PHANTOM VARIABLE			
7 (81)	X	X	
8 (97)	X	X	
9.5 (115)	X	X	
NO MEASURES—PHANTOM VARIABLE			
11.5 (140)	X	X	X
13 (157)	X	X	X
NO MEASURES—PHANTOM VARIABLE			
16.5 (198)	X	X	X

Note. *Age is rounded to the nearest half-year.
Phantom variables were used to improve estimation of the autoregression term in the latent growth model with structured residuals. They constrained the autoregression between adjacent waves with wide gaps between observations (e.g., 3.9 and 6.8 years) to be equal to the square of the autoregression between waves with narrower gaps (e.g., 6.8 and 8.1 years).

Missing Data

A number of baseline demographics and variables representing exposure to adversity were examined (see **Table D1** for a complete list) and included as saturated correlates (Collins, Schafer, & Kam, 2001) if they correlated $\geq |0.05|$ with either (1) any parcel score or, (2) with a dummy variable indicating that a parcel was missing. This procedure resulted in the inclusion of 34 auxiliary variables as saturated correlates.

Table D1. List of Auxiliary Variables Examined and Included in CFA Model Used To Generate Factor Scores

Predictor	Number of Relations with Outcome Variables (max = 23)	Included	Variable Type
Parent SES	23	yes	factor
Mother's education at gestation	23	yes	ordered category
Parents are home-owners	23	yes	binary
Marital status	23	yes	ordered category
Maternal psychopathology (8 months)	23	yes	binary
Maternal psychopathology (21 months)	23	yes	binary
Maternal psychopathology (33 months)	23	yes	binary
One adult in the household (47 months)	23	yes	binary
Neighborhood housing quality (21 months)	23	yes	binary
Neighborhood housing quality (33 months)	23	yes	binary
Financial score (8 months)	23	yes	binary
Financial score (21 months)	23	yes	binary
Financial score (33 months)	23	yes	binary
Partner cruelty (33 months)	23	yes	binary
ppregnum	22	yes	ordered category
One adult in the household (21 months)	22	yes	binary
One adult in the household (33 months)	22	yes	binary
Partner cruelty (47 months)	22	yes	binary
Mother's age at birth	21	yes	ordered category
One adult in the household (8 months)	21	yes	binary
Partner cruelty (21 months)	21	yes	binary
Either parent in trouble with law enforcement (21 months)	20	yes	binary
Male	14	yes	binary
Partner cruelty (8 months)	10	yes	binary
Child psychological or emotional abuse (42 months)	9	yes	binary
Child psychological or emotional abuse (30 months)	8	yes	binary
Race = White	7	yes	binary
Family instability (42 months)	7	yes	binary
Either parent in trouble with law enforcement (33 months)	5	yes	binary

Either parent in trouble with law enforcement (47 months)	4	yes	binary
preg_size	2	yes	binary
Either parent in trouble with law enforcement (8 months)	2	yes	binary
Family instability (18 months)	2	yes	binary
Family instability (30 months)	2	yes	binary
Child psychological or emotional abuse (18 months)	0	no	binary

Note. Auxiliary variables were included when correlation with at least 1 $r \geq |.05|$.

Table E1. Fit Statistics for Growth Models

Model	df	χ2	p	Satorra-Bentler Chi-square difference (k vs. k-1)	chi-square difference test	CFI	TLI	RMSEA	BIC	AIC
1. Linear Growth Model	28	11,334.80	< .0001	---	---	.815	.862	.186	-16,349	-16,401
2. Quadratic Growth Model	24	6196.57	< .0001	5278.72	< .0001	.899	.912	.149	-22,883	-22,964
3. Cubic Growth Model	19	3582.70	< .0001	2622.75	< .0001	.942	.936	.127	-26,214	-26,331
4. Quartic Growth Model	18	3543.09	< .0001	13.67	.0001	.942	.933	.130	-26,219	-26,344
5. Quartic Growth Model with Structured Residuals	**17**	**3422.53**	**< .0001**	**156.78**	**< .0001**	**.944**	**.931**	**.131**	**-26,482**	**-26,714**

Note. All quartic models constrained the quartic variance to zero because otherwise, the model was underidentified. The final model is bolded.

Table F1. Fit Statistics for Growth Mixture Models

Model	free parameters	Loglikelihood	BIC	aBIC	AIC	CAIC	AWE	adjusted LRT (k vs k-1)	cmPk	cmPA
Baseline Model[a]	18	13,375	-26,581	-26,638	-26,713	-26,563	-26,358	---	---	---
Mixture Models										
all growth factor variances fixed to zero, AR varies across classes										
2 class	15	11,587	-23,033	-23,081	-23,143	-23,018	-22,847	< .0001	0.00	---
3 class	22	12,449	-24,691	-24,761	-24,853	-24,669	-24,419	< .0001	0.00	---
4 class	29	12,819	-25,367	-25,459	-25,580	-25,338	-25,008	< .0001	0.00	---
5 class	36	13,201	-26,065	-26,179	-26,330	-26,029	-25,620	.0022	0.00	---
6 class	43	13,568	-26,734	-26,871	-27,051	-26,691	-26,203	< .0001	0.00	---
7 class	50	13,802	-27,135	-27,294	-27,503	-27,085	-26,517	.0022	0.00	---
8 class	57	14,031	-27,528	-27,709	-27,947	-27,471	-26,823	.022	1.00	0.00
Invariant variances, diagonal covariance, linear slope variance, AR varies across classes										
2 class	17	13,044	-25,929	-25,983	-26,054	-25,912	-25,719	< .0001	0.00	---
3 class	24	13,461	-26,697	-26,773	-26,874	-26,673	-26,400	.0004	0.00	---
4 class	31	13,886	-27,482	-27,580	-27,710	-27,451	-27,099	< .0001	0.00	---
5 class	38	14,293	-28,231	-28,351	-28,510	-28,193	-27,761	.0044	0.00	---
6 class	**45**	**14,739**	**-29,056**	**-29,199**	**-29,388**	**-29,011**	**-28,500**	**< .0001**	**0.00**	**1.00**
7 class	52	15,086	-29,684	-29,849	-30,067	-29,632	-29,041	.175	0.00	---
8 class	59	15,409	-30,265	-30,452	-30,699	-30,206	-29,535	.035	1.00	---

Invariant variances, non-diagonal covariance, linear slope variance, AR varies across classes

2 class	18	13150	-26,132	-26,189	-26,265	-26,114	-25,910	.0001	0.00	---
3 class	25	13594	-26,954	-27,033	-27,138	-26,929	-26,645	.0087	0.00	---
4 class	32	14025	-27,751	-27,853	-27,987	-27,719	-27,356	0.216	0.00	---
5 class	39	14497	-28,629	-28,753	-28,916	-28,590	-28,147	0.002	0.00	0.00
6 class	46	14814	-29,197	-29,343	-29,535	-29,151	-28,628	0.316	0.00	---
7 class	53	15167	-29,839	-30,007	-30,229	-29,786	-29,183	0.145	0.00	---
8 class	60	15487	-30,412	-30,603	-30,854	-30,352	-29,671	0.006	1.00	---

Invariant variances, diagonal covariance, linear slope variance, AR invariant across classes

2 class	16	13,001	-25,851	-25,902	-25,969	-25,835	-25,654	< .0001	0.00	---
3 class	22	13,331	-26,456	-26,526	-26,618	-26,434	-26,184	< .0001	0.00	---
4 class	28	13,660	-27,059	-27,148	-27,265	-27,031	-26,713	.0008	0.00	---
5 class	34	14,103	-27,887	-27,995	-28,137	-27,853	-27,466	.0028	0.00	---
6 class	40	14,368	-28,362	-28,489	-28,656	-28,322	-27,867	.127	0.00	0.00
7 class	46	14,646	-28,861	-29,008	-29,200	-28,815	-28,293	.036	0.00	---
8 class	52	14,888	-29,289	-29,454	-29,671	-29,237	-28,646	.004	1.00	---

Note. The final model selection is bolded. Boxes denote models selected by fit statistics within-class. The Bayes Factor was excluded because it was not informative in model selection (e.g., all values for k vs k+1 class models were 0). Mixture models with variances for quadratic slopes yielded unacceptable models (non-positive definite psi matrices) and were consequently excluded from the table.

a = The baseline model was parameterized in the following way: 1 class, quartic polynomial, with random linear and quadratic variances, higher order variances fixed to zero, 3 phantom variables, time-invariant autoregressive structured residual, residual variances constrained equal within 2 separate pieces: 47-128 months, 140-198 months.

AR = autoregressive parameter for structured residuals.

Table G1. Growth Parameters for the One-Class and Six-Class Models

Parameter	One-Class Model Estimate (SE)	Latent Class					
		Minimal Symptoms Estimate (SE)	Adolescent Spike Estimate (SE)	Late Childhood Peak Estimate (SE)	Childhood Decrease Estimate (SE)	High and Reversing Estimate (SE)	High and Renitent Estimate (SE)
Means							
Intercept	0.76 (0.003)	0.62 (0.006)	0.79 (0.02)	0.72 (0.03)	0.99 (0.02)	0.84 (0.03)	0.91 (0.02)
Slope	0.18 (0.02)	-0.21 (0.03)	-0.24 (0.15)	0.95 (0.16)	0.05 (0.16)	0.70 (0.13)	0.63 (0.06)
Quadratic	-0.21 (0.06)	0.59 (0.09)	1.89 (0.65)	-3.29 (0.66)	-2.08 (0.42)	-0.47 (0.53)	-0.89 (0.15)
Cubic	-0.07 (0.06)	-0.70 (0.10)	-3.18 (0.79)	4.75 (0.79)	2.05 (0.40)	-0.44 (0.64)	0.42 (0.015)
Quartic	0.09 (0.02)	0.26 (0.03)	1.47 (0.28)	-1.99 (0.28)	-0.53 (0.13)	0.30 (0.22)	-0.07 (0.05)
Structured Residual	0.30 (0.03)	0.43 (0.04)	0.49 (0.10)	-0.48 (0.06)	0.05 (0.07)	-0.50 (0.03)	0.90 (0.007)
Variances*							
Intercept	0.07 (0.001)	0.05 (0.002)	0.05 (0.002)	0.05 (0.002)	0.05 (0.002)	0.05 (0.002)	0.05 (0.002)
Slope	0.21 (0.028)	0.01 (0.001)	0.01 (0.001)	0.01 (0.001)	0.01 (0.001)	0.01 (0.001)	0.01 (0.001)
Quadratic	0.21 (0.091)	---	---	---	---	---	---
Cubic	0.04 (0.018)	---	---	---	---	---	---
Quartic	---	---	---	---	---	---	---

Note. *Variances were constrained equal across latent classes.

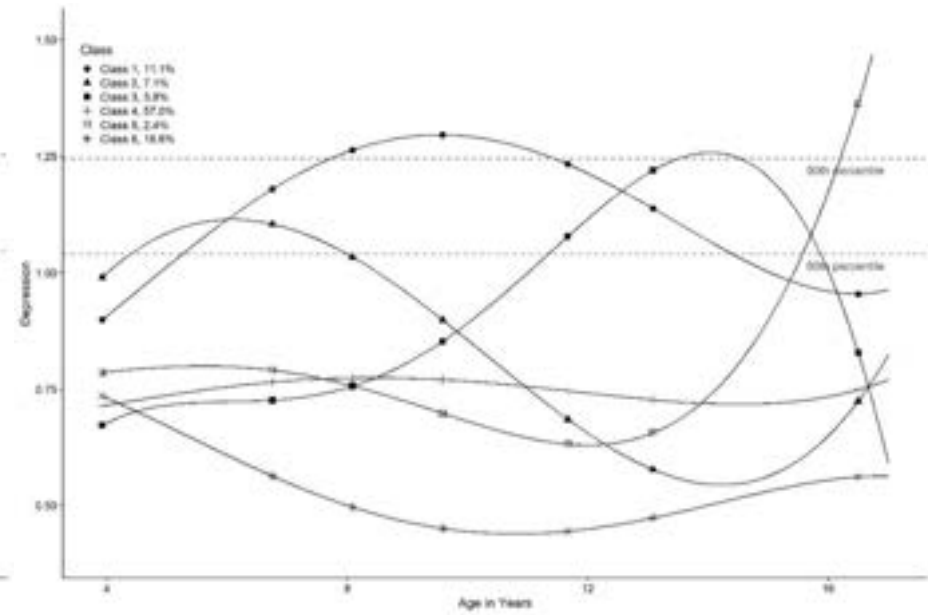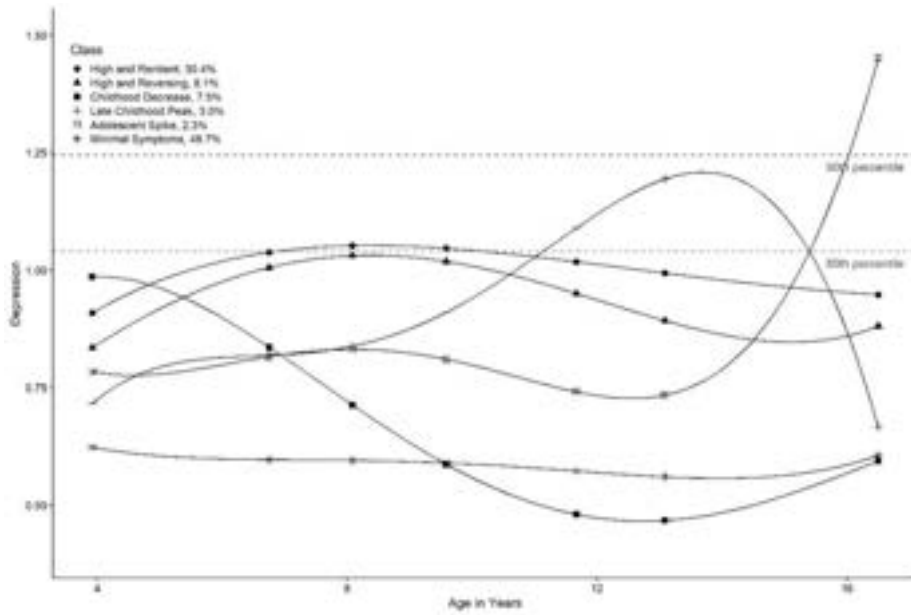**Figure H1.** Probability of Class Membership by Social Predictor Variables
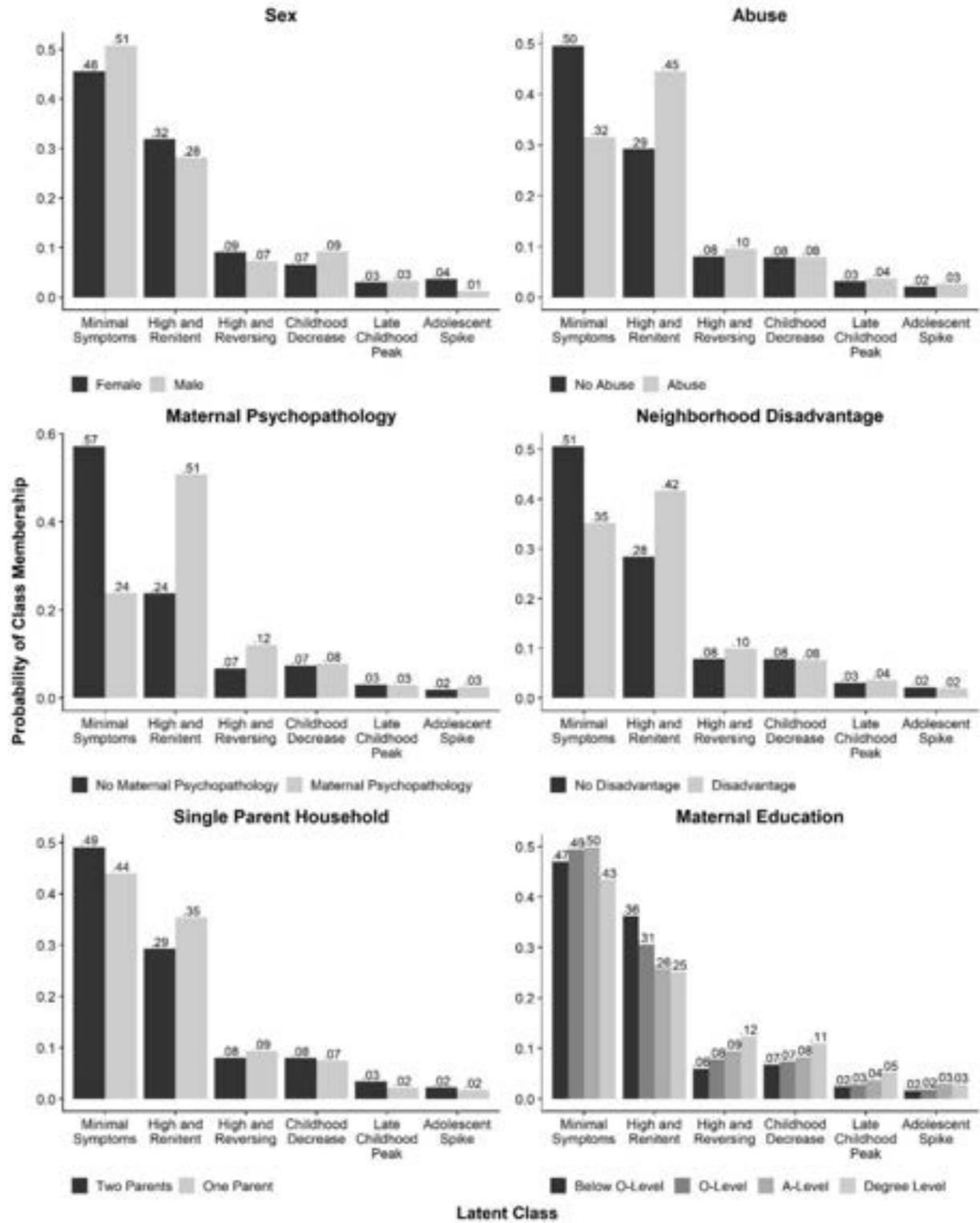
Note. For each level of each social factor, bars sum to 100%.

Figure I1. Latent Class Trajectories Using a Traditional Growth Mixture Model Without Structured Residuals

A. 6-Class Growth Mixture Model with Structured Residuals   B. 6-Class Traditional Growth Mixture Model

References for Supplement

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological methods*, *8*(3), 338.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233-255.

Muthen, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*(1-2), 43-65.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*(3), 354.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243-248.