Development and Psychopathology **19** (2007), 345–366
Copyright © 2007 Cambridge University Press
Printed in the United States of America
DOI: 10.1017/S0954579407070174

Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences

SONYA K. STERBA, MITCHELL J. PRINSTEIN, AND MARTHA J. COX
University of North Carolina at Chapel Hill

Abstract

Developmental psychopathology theory speaks to the existence of early-manifesting internalizing problems with a heterogeneous longitudinal course. However, the course of internalizing problems has been investigated largely from late childhood onward, with methods that assume children's problem trajectories vary more so in rate than in qualitative functional form. This can obscure heterogeneity in symptom process and course, obscure onset of early gender differences in internalizing problems, and obscure the relevance of early sociocontextual risks for long-term internalizing outcomes. The present study addressed these issues by using person-oriented (latent growth mixture) methods to model heterogeneity in maternal-reported internalizing symptoms from age 2 to 11 years ($N = 1{,}364$). Three latent trajectory classes were supported for each gender: two-thirds of children followed a low-stable trajectory; smaller proportions followed decreasing/increasing or elevated-stable trajectories. Although the number, shape, and predictive validity of internalizing trajectory classes were similar across gender, trajectory classes' initial values and rates of change varied significantly across gender, as did the impact of maternal postpartum depression and anxiety on latent growth factors. Extracted latent trajectories were differentially predicted by postpartum maternal psychopathology, and themselves, in several respects, differentially predicted self-reported depressive symptoms in preadolescence. However, discussion focuses on the need for further external validation of extracted latent classes.

Studies describing the longitudinal course of internalizing pathology have predominately focused on the middle childhood and adolescent period (e.g., Dekovic, Buist, & Reitz, 2004; Galambos, Barker, & Almeida, 2003; Kovacs et al., 1984), to the neglect of the preschool phase. This neglect may be reminiscent of historical biases that young children lack the sense and expectations of self to permit internalizing problems (Luby, 2000). This neglect may also stem from perceptions that internalizing problems markedly shift in presentation from preschool to preadolescence, precluding reliable measurement of construct change over time. However, this neglect does not reflect present knowledge that internalizing disorders do appear in preschoolers in typically manifesting forms (e.g., depression with neurovegitative features rather than mainly "masked depression"; Luby et al., 2003). Further, internalizing disorders appear in preschoolers at rates comparable to some externalizing disorders (Angold, Egger, Erkanli, & Keeler, 2005). Moreover, internalizing symptoms can show phenotypic consistency across childhood at molar and/or molecular levels (Mitchell, McCauley, Burke, & Moss, 1988; Muris, Merckelbach, Gadet, & Moulaert, 2000), although some shift in content or prevalence with age (see Carlson & Kashani, 1988). Omitting preschoolers from research on the longitudinal course of internalizing problems

The authors thank Daniel J. Bauer for his insightful comments on earlier versions of this manuscript.

Address correspondence and reprint requests to: Sonya K. Sterba, L. L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, Davie Hall, Campus Box 3270, Chapel Hill, NC 27599-3270; E-mail: ssterba@email.unc.edu.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

is thus not a methodological necessity. It is actually quite the reverse. Their omission could skew our perceptions of (a) homotypic continuity in internalizing problems, (b) emergence of early gender differences in internalizing pathology, and (c) relevance of early sociocontextual risks for long-term internalizing outcomes. In this study, we examine these three issues by tracking internalizing symptoms prospectively over a broader age span, toddlerhood to preadolescence.

Heterogeneity in Trajectories of Internalizing Symptoms

Longitudinal studies of childhood internalizing problems have also traditionally employed methods of modeling change that assume a common relationship between internalizing problems and time for *all* children (i.e., variable oriented methods; Beitchman, Wekerle, & Hood, 1987; Keenan, Shaw, Delliquadri, Giovannelli, & Walsh, 1998; Mesman & Koot, 2000). If this assumption is tenable, then average onset dates for mood disorders are readily interpretable (e.g., Costello et al. 1988; Kovacs et al., 1984), and we can speak of developmental trends in fears (e.g., in separation anxiety; Albano, Chorpita, & Barlow, 1996). If we *cannot* assume a common population trend for internalizing problems across childhood, a summary estimate of symptom change from a variable-oriented analysis (a) cannot be used to make inferences about how subgroups of children are changing, (b) may obscure patterns of change unique to subgroups, and (c) may have compromised interpretability (von Eye & Bergman, 2003). From a developmental psychopathology perspective, children whose environmental and genetic risk factors repeatedly interfere with resolution of socioemotional developmental milestones could exhibit symptom trajectories that vary—not in merely rate, but in *qualitative functional form*—from low-risk children's trajectories (Cicchetti & Toth, 1995; Raudenbush, 2001; Robins, 1966; Rutter, 1989). Moreover, changing contextual events and demands could yield equifinality or multifinality among some children's internalizing problem trajectories (Cicchetti & Rogosch, 1996). These complex change processes are not consistent with variable-oriented methods' assumption that variation around a mean internalizing trend is random.

Nevertheless, recent studies have continued to employ variable-oriented methods, even improving upon prior work by decoupling systematic from random change using conventional latent growth curve modeling. However, this recent work has yielded seemingly contradictory findings of maternal-reported childhood internalizing problems. For example, Colder, Mott, and Berman (2002) found that, on average, internalizing symptoms increase from age 4 to 8 years. Keiley, Lofthouse, Bates, Dodge, and Pettit (2003) found that symptoms are, on average, stable from 5 to 13 years. Gazelle and Ladd (2003) found that symptoms, on average, decrease from 6 to 10 years. It seems possible that in these samples there were unobserved subgroups of children following stable, increasing, and decreasing trajectories, but in each case the mean trajectory was biased in the direction of a certain prevalent pattern. This could be explicitly investigated by allowing latent classes of children with similar contributory processes to follow class-specific symptom trajectories, instead of restricting all children to follow a single mean trajectory. Such a (person-oriented) approach provides a more nuanced depiction of complex realities of developmental course than does a variable-oriented approach, facilitating examination of questions particularly relevant to developmental psychopathological theory (Cicchetti & Rogosch, 1996). Nonetheless, this approach still presents a vastly simplified depiction of the multiplicity of change processes occurring in nature.

The present study uses such a person-oriented method to chart heterogeneity in mother-reported internalizing trajectories from ages 2 to 11 years (i.e., latent growth mixture modeling [LGMM]; Múthen & Shedden, 1999; Nagin, 1999). We test whether the stability of internalizing problems often observed at the variable-level predominately reflects the many children with low symptoms across time, or whether an early-starter class and identify children who exhibit elevated problems from age 2.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

Specifically, most children are expected to exhibit low and stable internalizing symptoms, in line with Hollenstein, Granic, Stoolmiller, and Snyder (2004). A small proportion of children with elevated sociocontextual risk factors in infancy is expected to exhibit stable elevations in symptoms. This follows evidence that earlier emerging pathology is associated with intractable course, greater impairment, and difficulty resolving future developmental tasks (e.g., Barlow, 1988; Cicchetti & Toth, 1995). Another small proportion of children is expected to exhibit moderate-range symptoms that are unstable over time (e.g., Cantwell & Baker, 1989). These children may have a moderate accumulation of early risk factors that render them vulnerable to symptom fluctuations in response to environmental stresses and transitions (e.g., decreases in early childhood when separation fears remit). If evidenced, these results would represent an integration of prior findings (from Colder et al., 2002; Gazelle & Ladd, 2003; Keiley et al., 2003) into a more consistent depiction of the longitudinal course of childhood internalizing symptoms.

External Validation of Trajectories

One advantage of a growth mixture approach is that it allows estimation of class specific levels of predictors and outcomes. This simplifies interpretability and comparability of findings across different samples, as it eliminates the need for plotting interaction effects at ad hoc selections of values (which *may or may not* be representative of actual patterns in the data; Bauer & Shanahan, in press[1]). This also allows us to test the external validity of extracted latent classes. Specifically, we examine whether children's membership in an internalizing trajectory class can be predicted from and predict variables *other* than those used to create the classes (von Eye & Bergman, 2003) in ways that would "not necessarily be expected if the population was homogeneous" (Bauer & Curran, 2003, p. 359).[2] This helps determine whether these classes correspond with qualitatively different underlying subpopulations of children, or whether classes should more readily be viewed as demarcating clinically relevant thresholds on a complex distribution of internalizing symptoms (Múthen, 2004). To this same end, we also test whether the maternal-reported internalizing problem trajectory a child follows from age 2 to 11 years is predictive of their *self-reported* internalizing symptoms as a preadolescent.

The predictor variables of interest in this study are those that can help explain heterogeneity in the course of internalizing problems from toddlerhood onward. Specifically, we focus on risk factors that could interfere with the toddler/preschool-age "sensitive period" for the organization of emotion regulation and cognitive competencies and formulation of self-other relations (see Cicchetti, Rogosch, & Toth, 2000). These factors could potentiate internalizing problems at this period or subsequently. Maternal internalizing psychopathology during the postpartum and infancy period is one such risk factor that can predict preschoolers' internalizing symptoms (e.g., Philipps & O'Hara, 1991) even at subsyndromal levels (Field et al., 1996). A broad array of interrelated mechanisms have been implicated in this intergenerational transmission, including genetic predispositions, impaired parenting sensitivity and consistency, and emotional socialization/modeling of maternal behavior (see McLennan & Offord, 2002). In this study we focus solely on the latter, as it has been shown to be relevant both for the transmission of anxious and depressive behavior (Hsu, 1996; Murray & Cooper, 1997).

Maternal psychological unavailability/detachment (Egeland & Sroufe, 1981) and maternal negative thoughts and affect (Campbell, Cohn, Flanagan, Popper, & Meyers, 1992) may directly foster a depressotypic organization in children, with possible mechanisms including children mirroring mothers' negative affect.

1. Bauer and Shanahan (in press) also noted that LGMM does not assume particular functional forms for predictive relationships or interactions (e.g., linearity).

2. For example, merely finding that high, medium, and low classes correspond with high medium and low levels of an outcome is not enough to rule out the notion of a single population distribution.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

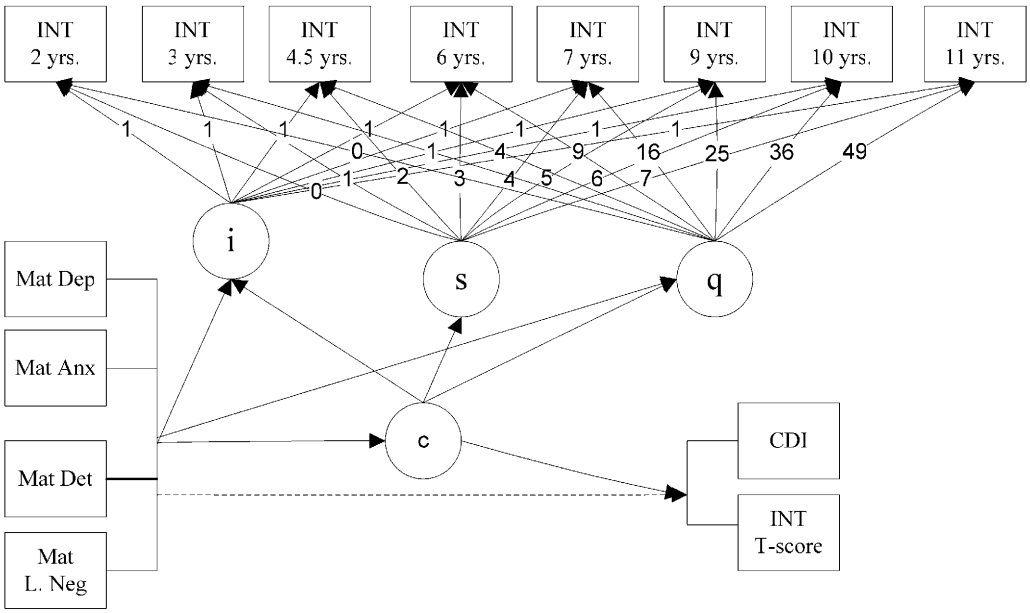


Figure 1. A path diagram of the latent growth mixture model (LGMM; estimated each for boys and girls). INT 2–11 years, child internalizing symptoms ages 2–11 years; *i*, latent intercept factor; *s*, latent linear slope factor; *q*, latent quadratic slope factor; *c*, latent classification variable; Mat Dep, maternal depressive symptoms; Mat Anx, maternal separation anxiety symptoms; Mat Det, maternal detachment; Mat L. Neg, maternal lack of negativity; CDI, CDI clinical cutoff; INT *T* score, CBCL internalizing *T*-score clinical cutoff. Each latent growth factor (*i*, *s*, *q*) is measured by the eight observed internalizing scores, with loadings fixed according to Singer and Willett (2003). Not shown because of space limitations are residual terms for all endogenous variables, variances for all exogenous variables, and covariances of growth factor residuals within class for all LGMMs.

For example, mothers with postpartum internalizing disorders have infants who, in controlled experiments, exhibit more negative contingent responses and fewer positive responses (Cohn, Campbell, Matias, & Hopkins, 1990). These infants are physiologically hyperaroused in interactions with their mothers, while exhibiting passive and detached behavior, which generalizes to interactions with strangers (Field et al., 1988). In addition, mothers showing psychological unavailability or depression postnatally can have infants with lower cognitive functioning in infancy and/or preschool (Egeland & Sroufe, 1981; Murray & Cooper, 1997).

However, the majority of research on the predictive role of maternal postpartum psychopathology looks only at infant and preschool outcomes. Yet, the impact of early maternal psychopathology could be longstanding if children's resultant emotional/cognitive disorganization in preschool impedes adaptive organization at future sensitive periods. In the present study, we investigate whether mothers' *own* postpartum internalizing symptoms intergenerationally predict children's entire internalizing trajectory from age 2 to 11 years. We hypothesize that mothers' higher observer-reported symptoms (detachment and negativity) and self-reported symptoms (depression and separation anxiety) will positively affect the intercept and slope of children's internalizing symptoms. To test the external validity of extracted latent internalizing classes, we investigate whether mothers' symptoms will also predict children's class membership in an increasing or stably elevated trajectory class, over a low trajectory class (shown in Figure 1). The latter finding would suggest the usefulness of classification, and the relevance of early maternal psychopathology as a etiological factor in the longitudinal sequellae of

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

child internalizing, through preadolescence, not merely preschool.

Gender Differences in Internalizing Symptom Trajectories

In addition to depicting heterogeneity in patterns of symptom change and externally validating these patterns, the third goal of this study was to examine how these longitudinal patterns of internalizing problems and class-specific risk factors vary across gender. Although we view person-oriented methods as most theoretically consistent for modeling internalizing symptoms, we nonetheless can examine children's symptom trajectories (number, shape, covariates, and growth factors) *across* gender. This amounts to integrating both person- and variable-oriented analyses. This is a much-needed step, as the nature and timing of gender differences in childhood internalizing problems has been theoretically contested but less subject to comparably rigorous examination. Moving beyond traditional methods of cross-gender comparisons of symptom prevalence at discrete points in time, the present study translates opposing theoretical positions into specific hypotheses about behavioral change.

Some theorists assert that both boys and girls similarly experience internalizing problems across early/middle childhood (e.g., Keenan & Shaw, 1997; Rudolph, 2002) with prominent gender differences not emerging until adolescence (Cicchetti & Toth, 1995; Zahn-Waxler, Klimes-Dougan, & Slattery, 2000) or preadolescence (Costello et al., 1988; Muris et al., 2000). This theory translates into our hypothesis that equivalent proportions of boys and girls may follow similar trajectory types throughout early childhood, yet in preadolescence develop disparities, possibly in the form of an increasing trajectory for girls not boys. Others contend that gender differences emerge early, with girls showing less internalizing and externalizing pathology in early childhood as a result of their relatively greater socioemotional strengths to meet early developmental challenges (Sommers, 2000). This theory translates into our counterhypothesis that fewer girls than boys may follow high and stable internalizing trajectories, and that girls' stably elevated trajectory classes may have lower amplitudes than their counterpart boys' classes.

At least two previous studies have compared internalizing trajectories across gender, with sufficient sample sizes, separating out measurement error. Both found that the slopes and intercepts of the internalizing trajectories for boys and girls were similar, yet temperamental and peer covariates of these growth factors varied across gender (Brendgen, Vitaro, Bukowski, Doyle, & Markiewicz, 2001; Colder et al., 2002). Although valuable, these studies have several limitations. First, Colder and colleagues (2002) restricted growth to following a single curve, so it may be that gender differences occur between slopes and intercepts of specific trajectory classes of children, but are obscured when comparing the entire mean trajectories across gender. Second, Brendgen and colleagues (2001) made descriptive rather than inferential cross-gender comparisons. The present study performs the more rigorous inferential testing of the equality of each classes' covariates, slope, and intercept across gender by entering boys' and girls' separate latent growth mixture models into a multiple-group framework. It is unlikely that gender differences are so stark as to be entirely absent in one age range (childhood) and emerge suddenly with prominence in another (adolescence). This again motivates our third aim: testing whether a detailed investigation of gender differences over an underinvestigated age span will reveal nuanced insights regarding timing and type of gender differences in internalizing problems.

Method

Participants

Data from children in the National Institute of Child Health and Development (NICHD) Study of Early Child Care and Youth Development were utilized in this study. Recruitment and selection procedures are described elsewhere (e.g., NICHD ECCRN, 2004a, 2004b) but outlined here. Families were recruited through hospital visits shortly after the birth of a child at 10 sites in the United States

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

(Little Rock, AR; Irvine, CA; Lawrence, KS; Boston, MA; Philadelphia, PA; Pittsburgh, PA; Charlottesville, VA; Morganton, NC; Seattle, WA; Madison, WI). All women giving birth during selected 24-hr sampling blocks were screened for eligibility based on criteria that included maternal age of at least 18, maternal ability to speak English, infant not one of a multiple birth, not released for adoption, and not evidencing congenital abnormalities.

Of the 8,986 total mothers who gave birth during the sampling intervals, 5,416 (60%) met eligibility criteria and gave permission to be called 2 weeks later and 3,015 (56%) were conditionally randomly sampled for the call (intended to permit self-weighting). Conditioning ensured adequate representation (at least 10%) of each single mothers, ethnic minority mothers, and mothers without a high school degree. A total of 1,526 met additional eligibility criteria at the call and agreed to an interview; 1,364 of these became the study participants by completing an interview at 1 month. The resulting sample was diverse: 24% of infants were ethnic minority (Asian or Pacific Islander, $N = 22$; Black or African American, $N = 176$; other, including Hispanic, $N = 64$; American Indian, Eskimo, Aleut, $N = 5$; and White, $N = 1{,}097$); 48% of infants were female; 11% of mothers had not completed high school, and 14% were single mothers. The 1,364 participating families were similar in maternal education, ethnic minority composition, and rate of single parenthood to the eligible hospital sample of 5,146 and similar to families living in the same census tracts nationally on key demographic variables (NICHD ECCRN, 2004a). This sampling design is fortunate for the present investigation. Prior studies using similar methodology have often relied on clinical or high-risk samples (e.g., Broidy et al., 2003; Gilliom & Shaw, 2004). This sample allows exploration of whether similar trajectory patterns generalize to a community sample, yet also features variability at upper extremities of the internalizing spectrum helpful for discriminating low base rate chronic classes (Nagin, 1999).

Missing data on observed outcome variables showed no statistically reliable deviation from randomness using Little's MCAR test ($p = .73$). Thus, analyses utilized robust full information maximum likelihood estimation, enabling use of all available data and producing unbiased parameter estimates and standard errors. As Mplus 3.12 software does not allow missing data on observed covariates (outcomes are modeled conditional on them and they have no distributional assumption), in analyses involving maternal covariates, the 128 cases missing covariate data and 21 cases with covariate data but missing all observed outcomes were listwise deleted (LD). Multiple imputation (MI) was not used for these cases as no available software imputes data from a complete data model where different parameters hold for latent subgroups of the population, as in our multiple class mixture analysis (see Múthen & Asparouhov, 2002). Under MCAR (which could not be rejected), LD estimates are consistent for first and second moments of the data, whereas MI is consistent only in the first moments. Sample size following LD allowed sufficient power for all growth modeling procedures, and even for estimating separate models by gender (see Eggleston, Laub, & Sampson, 2004, for sample size requirements). There were no significant differences between cases with and without missing data in means of maternal covariates (maternal depression, $M = 11.33$ vs. 11.36; maternal anxiety, $M = 69.71$ vs. 70.34; maternal lack of negativity, $M = 5.17$ vs. 5.03; maternal detachment, $M = .08$ vs. .09) using a Satterthwaite approximation. Variances and covariances of maternal covariates were also compared for cases with and without missing. Bartlett's modified likelihood ratio test of equality of covariances did not indicate a significant departure from homogeneity, $\chi^2\,(10) = 9.20$, $p = .51$. The folded form of the F statistic was used to test equality of variances, and there was a significant departure from homogeneity of variance only for lack of negativity, $F\,(1084, 148) = 1.45$, $p = .005$.

Procedure

Demographic information and multiple measures of child, family, and peer functioning were collected prospectively, in multiple settings, when children were 1, 6, and 15 months

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

and approximately 2, 3, 4.5, 6, 7, 9, 10, and 11 years old.

Measures

Proximal outcome.

Child internalizing symptoms (mother report). The NICHD Study of Early Child Care used mother reports of the 99-item Child Behavior Checklist (CBCL)/2–3 (Achenbach, 1992) to assess problem behaviors for the 2- and 3-year time points, and the 113-item CBCL/4–18 (Achenbach, 1991) at the 4.5-, 6-, 7-, 9-, 10-, and 11-year time points. On both versions, mothers rated how well each item described the target child on a 3-point scale. CBCL scales are empirically derived yet significantly relate to *DSM-III-R* classifications of internalizing disorders for preschoolers (Keenan, Shaw, Walsh, Delliquadri, & Giovannelli, 1997), and school-aged children (Jensen & Watanabe, 1999).

Modeling developmental pathways of internalizing symptoms over time requires identifying a subset of internalizing symptoms that are developmentally appropriate and phenotypically expressed from preschool to preadolescence (Singer & Willett, 2003). Although heterotypic continuity is common across childhood both at the disorder level (e.g., changes from externalizing to internalizing, and internalizing to externalizing disorders), and at the symptom level (e.g., depression with irritability vs. depression with neurovegetative features) studies have identified homotypic continuity for some internalizing symptoms over time (e.g., Birmaher et al., 2004; Muris et al., 2000). We model change over time in a subset of internalizing symptoms known to have greater developmental invariance (Achenbach, 1991, 1992). Of the 59 items that appear on *both* the CBCL 2–3 and CBCL 4–18, 14 items fall into the internalizing domain on at least one form. Of those 14, 3 items switch from internalizing to social problems across forms ("acts too young for age," "clings to adults or too dependent," "doesn't get along with other kids") and 3 more switch from internalizing to externalizing problems across forms ("doesn't feel guilty after misbehaving," "demands a lot of attention," "stubborn, sullen or irritable"). This leaves a total of eight items ("underactive, slow-moving, or lacks energy"; "unhappy, sad, or depressed"; "withdrawn, doesn't get involved with others"; "overtired"; "nervous, highstrung, or tense"; "too fearful or anxious"; "shy or timid"; "self-conscious or easily embarrassed") empirically operationalized by Achenbach (1991, 1992) to be homotypically internalizing across forms, from ages 2 to 18 years. The score on these eight items is used as the dependent variable in the current study (paralleling procedures of NICHD ECCRN, 2004b, for externalizing symptoms). These eight form-invariant items showed high internal consistency at all eight time points. Alphas ranged from .84 to .88 for boys and .85 to .87 for girls, suggesting that the selected item subsets represent cohesive constructs.

To heighten interpretability of class trajectories that are high, moderate, or low relative to other classes for that gender, the trajectories based on these eight items are used to predict a distal outcome of mother-reported preadolescent internalizing CBCL *T* scores. Growth mixture models thus output the probability of scoring above the CBCL internalizing clinical cutoff ($\geq$67) at age 10 for each latent class.

Distal outcome.

Preadolescent depressive symptoms (self-report). Preadolescents completed the 10-item Children's Depression Inventory (CDI) Short Form (Kovacs, 1992) assessing their level of depression over the past 2 weeks, when they were approximately 11 years old. The CDI Short Form uses the 10 most discriminating and internally consistent items from the 27-item form ($\alpha = .80$). A proportionally weighted sum of the items (rescaled from 1–3 to 0–2) yielded a score with a possible range of 0–20. As Mplus 3.12 only can handle dichotomized distal outcomes of trajectory class membership,[3] CDI Short Form scores were

3. Software capable of modeling continuous distal outcomes would have been advantageous; however, Mplus 3.12 was unique among GMM software in allowing us to perform a multiple-group GMM analysis. Initial GM models were run both in SAS Proc Traj (Nagin, 1999) and Mplus.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

dichotomized at scores of 8 to indicate "screening above average." This is the level NICHD ECCRN has applied for girls, generalized to both genders in light of Matthey and Petrovski (2002). They showed that Kovacs' (1992) original cutoff scores are based on a flawed base rate calculation that misses 86% of depressed children in a general population sample. The CDI has been shown to more validly discriminate between clinical and normal populations than between specific internalizing disorders (e.g., major depression and dysthymia; see Kovacs, 1992, p. 39), which complements how the CDI-S is used in the present study. The concurrent validity of the CDI is adequate, as it significantly correlates with theoretically related constructs such as lower self-esteem, self-concept, and social adjustment (Kovacs, 1992).

Time-invariant covariates.

Maternal separation anxiety symptoms (self-report). At the 1-month postbirth home visit, mothers filled out a 21-item scale constituting the maternal separation anxiety factor from Hock, Gnezda, and McBride's (1983) Maternal Separation Anxiety Scale. This scale assesses mothers' level of worry, sadness, and guilt when separated from her infant, and her beliefs about whether her child prefers and is better off in her care, and whether her child could adapt to nonmaternal care. The scale utilizes a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). The internal consistency for this maternal separation anxiety factor was $\alpha = .71$, with a test–retest reliability of $r = .71$. The coefficient of congruence comparing the factor structure at two times, approximately 3 months apart, was 1.0, with a possible range of 0 to 1 (Hock, DeMeis, & McBride, 1987).

Maternal depressive symptoms (self-report). At the 1-month postbirth home visit, mothers filled out the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977), which assesses symptoms of depression in nonclinical populations. Mothers rated their frequency of experiencing 20 symptoms during the past week on a 4-point scale (1 = *rarely or none of the time* [<1 day] and 4 = *most or all of the time* [5–7 days]). Rescored from 1–4 to 0–3, probability-weighted summed scores had a theoretical range from 0 to 60. The CES-D evidenced an internal consistency of $\alpha = .85$ for a general population sample ($N = 2{,}514$), high correlations with other measures of depressive symptoms, and adequate discriminative validity between psychiatric inpatients and general population adults (70 vs. 21% exceeding clinical cutoffs; Radloff, 1977).

Maternal detachment (observer report). At 6-months postbirth, mothers and infants participated in a 15-min semistructured play interaction in their homes. The procedure was designed to elicit naturalistic mother–child interactions highlighting a mother's capacity to interact in a sensitive, warm, and stimulating manner with her infant. Two trained coders rated maternal behavior on domains developed for the NICHD ECCRN including detachment. Detachment was operationalized as the degree to which the mother is emotionally and physically uninvolved with the child and unaware of the child's needs for interaction to facilitate involvement with objects or people. Detachment was rated on a 4-point scale (1 = *not characteristic*, 4 = *highly characteristic*), but was subsequently dichotomized. Winer reliability estimates were adequate ($r = .65$), as derived from a subsample ($N = 218$) of interactions rated by both coders.

Maternal lack of negativity (observer report). During the course of the 15 months' postbirth home visit, research assistants collected observational and interview information to complete the Infant Toddler–Home Observation for Measurement of the Environment (IT-HOME). The IT-HOME assesses the quality and quantity of stimulation and support available in the home for children aged 0–3 years, and shows adequate test–retest reliability from 12 to 24 months ($r = .77$; Bradley, 1994; Caldwell & Bradley, 1984). Of the 45 binary (yes/no) items, only those three pertaining to acceptance/lack of maternal negativity subscale were used for the present study. These included lack of observed or reported

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

maternal derogation, restraint, and yelling at child.

Analytic strategy

Growth/growth mixture portion. A latent variable framework was used, which accounts for dependency of repeated measures and for measurement error in endogenous variables, among other advantages. As a comparison point, conventional single-class latent growth modeling (LGM; see Múthen & Múthen, 2002, for review) was first employed in Mplus 3.12 (Múthen & Múthen, 1998–2004) to model internalizing symptoms at eight time points (2, 3, 4.5, 6, 7, 9, 10, and 11 years). LGM assumes growth trajectories in the sample arise from a single distribution. As such, this method captures individual differences in internalizing behavior as random effects around common latent growth factors. A latent intercept factor and latent linear slope factor (and where supported, a latent quadratic slope factor) were measured by the observed internalizing scores at the eight time points. Latent intercept and slope factors were predicted by four maternal variables,[4] and these latent factors themselves predicted distal outcomes of preadolescent CDI scores and CBCL *T* scores. Separate models were estimated for girls and boys. A path diagram of this model is found in Figure 1, specifics of which are discussed below.

Second, person-oriented LGMM (Múthen & Shedden, 1999) was employed. LGMM is an approach that relaxes the single population assumption of LGM. Instead of considering individual variation around a single mean curve, LGMM allows different, latent, classes of individuals to vary around different mean curves (Bauer & Curran, 2003). This is accomplished through the addition of a latent categorical classification variable (shown as *c* in Figure 1), which influences the internalizing growth factors. Because classification is accomplished using model-based, probabilistic assignment of children to latent classes, it is not subject to the typical critiques of heuristic algorithms (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002) used to categorize individuals into disjoint groups based on a priori grounds. As growth is summarized by the latent classification variable, it naturally was allowed to predict CDI and CBCL *T*-score distal outcomes directly via a logistic regression. In the LGMM models, Figure 1 shows that maternal covariates were allowed to (a) influence slope and intercept growth factors directly, and indirectly through the latent classification variable, as well as (b) influence preadolescent CDI and CBCL *T*-score scores directly. Effects of each maternal covariate on class membership were assessed conditional on the others in the model.[5] If one notes that the LGM is a special case of the LGMM in which $c = 1$, Figure 1 is seen to represent both the LGM and LGMM models. In the NICHD ECCYD, children were nested within site as well as within time. However, including site as a third level of nesting was not warranted as the proportion of variance due exclusively to differences between sites was very small (interclass correlation = .01; see Hox & Maas, 2001).[6]

The best-fitting models for boys and for girls from the above separate LGMM analyses can be compared descriptively in terms of

4. Additional demographic background predictors were not included because our aims were mainly descriptive in nature. If our focus had been on prediction (e.g., to use classes as the basis for treatment assignment), we would have conditioned class probabilities on other demographic factors. Nevertheless, supplemental analyses suggested that our omission did not alter substantive findings. Effects of postnatal income to needs ratio, for example, did not significantly predict growth factors directly (or indirectly through class membership) when included with the set of maternal covariates.

5. To maintain model identification, one class was designated the reference class, and comparisons were conducted against the other classes.

6. Moreover, with only 10 sites, it would have been infeasible to incorporate site as a random effect, as there would have been more parameters than clusters (see Hox & Maas, 2001). In addition, Mplus 3.12 does not permit three levels of nesting in models with two categorical latent variables. The fixed effects in our model are marginal estimates pooling over site, and these are unbiased because our estimation method (full information maximum likelihood estimation) incorporates pseudomaximum likelihood estimates and yields Huber–White corrected standard errors for clustered or nonclustered data.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

(a) number of latent classes; (b) slope, intercept, and proportion of boys and girls in each class; (c) covariate influence on latent class type; and (d) classes' predictive validity for distal outcomes. However, to facilitate formal inferential testing of cross-gender differences on these four fronts, both best-fitting models were combined into a multiple-group growth mixture model.

Multiple group portion. Cross-gender invariance was tested in stages within a multiple-group latent variable framework. Successively more stringent tests of cross-gender invariance were performed at each stage, contingent on the establishment of invariance at the prior stage. First, *configural* invariance was tested (i.e., whether both groups (genders) had the same model specification and whether the multiple group model showed acceptable fit when both models were estimated simultaneously without further constraints; Meredith, 1993). In our case, this means that both genders needed to independently evidence the same number of latent classes (e.g., three), and same type and number of latent growth factors (e.g., an intercept, linear slope, and quadratic slope) during growth mixture estimation, for meaningful multiple-group invariance testing to commence (L. K. Múthen, personal communication, May 23, 2005). If configural invariance held, *weak* invariance of factor loadings across was then tested. Because both the boys' and girls' models were growth models with the same types of growth factors, the measurement portion of each model already had loadings on each factor fixed at the same values across gender (see Figure 1). Thus, if configural invariance held for our growth models, weak invariance automatically held as well. If both held, *strong* invariance was then tested, where factor means, variances, and covariances were compared across groups. Subsequently, path coefficients for the nonmeasurement portions of the model were compared. Yet, in contrast to traditional procedures for testing strong invariance (Meredith, 1993), our multiple group GMM was unique in having not just one mean and variance for each factor for boys and one mean and variance for each factor for girls. Rather, the mean and variance for *each* class on *each* factor was compared across gender. The factor covariances and path coefficients for predictor variables were then also compared (all with nested chi-square tests). This third stage of cross-gender invariance testing is focused on in the Results section.

Model selection and model fit. In LGMM, model comparison does not begin with fitting an unconditional model, as excluded effects of covariates on growth factors can result in incorrect class probability estimates (Múthen, 2004, p. 354). All model comparisons were performed conditional on covariates of trajectory classes and distal outcome of trajectory classes. Initially, both a latent growth mixture model (allowing residual within-class variation and covariation of growth factors) and the more parsimonious latent class growth model (constraining within-class residual growth factor variances and covariances to 0) were estimated. If not significantly different, the LCGM was used. All models were estimated with multiple random sets of starting values to prevent convergence on a local maxima.

Multiple measures of model fit were considered when comparing models. Smaller relative values of information criteria were sought, which included the Bayesian information criteria (BIC) and Akaike IC (AIC), both weighting goodness of fit to the data against model parsimony. The Lo, Mendell, and Rubin (2001) likelihood ratio-based chi-square test statistic (LMR-LRT) was used to compare the absolute fit of a m-class model with an $m + 1$ class model, and a low p value indicates that the data are more likely to be generated by the specified model than by one with one less class.

Classification accuracy. As membership in trajectory classes was unobserved, each child was assigned to the class for which they had the highest posterior probability of class membership (Nagin, 1999). Precision of class assignment was indicated by mean assignment probabilities for each class ($>.70$ is considered good, 1.0 represents complete certainty of assignment; Nagin, 1999). A standardized

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

summary measure of classification accuracy, entropy, was also employed.

Results

Descriptive statistics

Descriptive statistics for all study variables are shown in Table 1. On average, boys and girls exhibited low levels of internalizing problems that declined over time. Postpartum maternal detachment was rare, and most mothers lacked negativity at 6 months postbirth. Yet some elevation in postpartum maternal depressive or anxiety symptoms was normative and correlated with child internalizing scores at later time points, albeit more strongly at younger ages.

Conventional versus mixture models of internalizing symptoms

We first examined whether changes in internalizing symptoms were well described by variation around a single mean trajectory (LGM; conventional growth model), or whether allowing class-specific developmental trajectories (LGMM) yields a more statistically fitting and substantively compelling depiction of change. As articulated previously, model selection was guided by seeking smaller information criteria values, lower entropy, and larger log-likelihood values. Best-fitting models for both genders allowed within-class residual variation, as constraining such variation to 0 within class significantly reduced model fit for boys, χ^2 (6, 646) = 415.18, $p < .001$, and for girls χ^2 (3, 577) = 165.0, $p < .001$. Neither the boys' nor girls' best-fitting models showed a detectable improvement in model fit from inclusion of direct effects of maternal covariates on distal outcomes (dashed line in Figure 1), so these paths were not included in further models, χ^2 (10, 646) = 11.01, *ns*, for boys, and χ^2 (10, 577) = 2.528, *ns*, for girls. Best-fitting three-class models for both genders included one class with a linear slope only and two classes with both linear and quadratic slope factors. In each model, the residuals of the latent growth factors were significantly intercorrelated. Table 2 compares a series of models fit for girls and boys. Table 3 provides class-specific intercepts, slopes, and mean posterior probabilities for final best-fitting girls' and boys' models selected from Table 2.

Boys' internalizing. For boys, the BIC, AIC, log likelihood, LMR-LRT, and entropy criteria all found that a two or more class (LGMM) model provided a better fit to the data than the single class (LGM) model. The next step was to compare two-, three-, and four-class LGMM models. The BIC favored the three-class model, and the LMR-LRT as well preferred a three- to a two-class model. We could not estimate the LMR-LRT for the four-class model with a linear and quadratic slope, but as an approximate check, the LMR-LRT from a four-class model with only a linear slope was not significant ($p = .84$). The four-class model yielded substantively less relevant information, as it merely split one low trajectory class into two. Although entropy for the three-class model is slightly lower than the two- and four-class models, Table 3 shows that mean posterior probabilities for each class of the boys' three-class model well exceeded adequacy standards. The log likelihood, lacking a penalty for addition of parameters, continued to increase for each model. A five-class model did not converge, indicating further model comparison was not supported (Nagin, 1999). Thus, a three-class model was accepted as best fitting.

Initial internalizing scores for each class of the boys' best-fitting model were significantly different than 0, as shown in Figure 2 and in Table 3. The nearly one fifth of boys comprising Class 1 showed *elevated-stable* internalizing symptoms across childhood ($N = 77$, 13%). This class showed a nonsignificant linear slope. Another fifth of boys, in Class 2, and had *decreasing/increasing* internalizing symptoms that declined from ages 2 to 6 and then rose until age 11 ($N = 135$, 22%). This class had a significant negative linear slope and significant positive quadratic slope. Finally, the majority of boys in Class 3, had mother-reported internalizing symptoms that remained *low* across childhood, declining from low to near zero ($N = 386$, 65%). This class had a significant negative linear slope and small but

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

Table 1. *Descriptive statistics*

	M. Det	M.L. Neg.	M. Dep.	M. Anx.	CDI	CBCL Int-T	Int 2	Int 3	Int 4.5	Int 6	Int 7	Int 9	Int 10	Int 11	*M*	*SD*	*N*	Range
M. det.	—	−.14**	.05	.11**	−.04	−.07	.04	.01	−.06	−.07	−.03	−.06	−.07	−.04	0.08	0.27	614	0–1
M.L. neg.	−.08*	—	−.17**	−.21**	−.06	−.02	−.16**	−.16**	−.04	−.05	−.07	.01	.00	.06	5.19	1.03	601	0–6
M. dep.	.06	−.09*	—	.21**	.08	.26**	.23**	.25**	.20**	.19**	.10*	.14**	.18**	.16**	10.95	8.60	658	0–46
M. anx.	.09*	−.11**	.22**	—	.04	.06	.12**	.13**	.04	.01	.07	.04	.03	−.01	70.48	13.34	652	37–105
CDI	.00	.01	.04	.05	—	.16**	.06	.08	.06	.03	.10*	.11**	.18**	.13**	0.02	0.13	511	0–1
CBCL-T	.01	−.04	.21**	.05	.02	—	.38**	.37**	.43**	.43**	.51**	.59**	.83**	.62**	47.95	9.22	506	33–81
Int 2	.01	−.06	.18**	.13**	.10*	.28**	—	.55**	.47**	.40**	.40**	.36**	.34**	.35**	3.06	2.00	578	0–11
Int 3	.04	−.12**	.24**	.15**	.00	.32**	.58**	—	.52**	.47**	.44**	.41**	.33**	.35**	3.34	2.25	571	0–14
Int 4.5	.00	−.02	.14**	.06	−.09	.39**	.32**	.49**	—	.61**	.55**	.47**	.44**	.46**	1.81	1.71	532	0–10
Int 6	−.02	−.07	.12**	.03	−.01	.49**	.26**	.42**	.57**	—	.57**	.54**	.42**	.42**	1.54	1.63	525	0–8
Int 7	.00	−.03	.19**	.00	−.02	.54**	.31**	.39**	.59**	.64**	—	.58**	.51**	.50**	1.65	1.52	513	0–8
Int 9	−.01	.00	.13**	.01	.03	.60**	.23**	.29**	.47**	.52**	.60**	—	.62**	.55**	1.68	1.81	519	0–8
Int 10	−.04	.01	.16**	.03	.03	.83**	.27**	.30**	.43**	.52**	.63**	.71**	—	.65**	1.61	1.68	506	0–10
Int 11	−.06	−.04	.16**	.08	.04	.65**	.20**	.33**	.45**	.53**	.62**	.64**	.72**	—	1.78	1.82	512	0–10
M	0.10	4.91	11.75	70.05	0.02	47.77	3.04	3.21	1.73	1.48	1.61	1.64	1.55	1.70				
SD	0.30	1.17	9.38	13.18	0.14	10.13	2.00	2.06	1.63	1.64	1.76	1.91	1.83	1.95				
N	658	633	705	699	508	516	612	605	542	533	515	507	516	508				
Range	0–1	0–6	0–53	35–105	0–1	33–82	0–11	0–14	0–10	0–10	0–13	0–16	0–13	0–13				

Note: Data for girls are shown above the diagonal, and data for boys are below the diagonal. M. det., maternal detachment; M.L. neg., maternal lack of negativity; M. dep., maternal depressive symptoms; M. anx., maternal separation anxiety symptoms; CDI, Child Depression Inventory clinical cutoff; CBCL Int-T, CBCL internalizing *T* score; Int 2–11, internalizing symptoms at ages 2–11 years.
*$p < .05$. **$p < .01$.

Table 2. *Model selection and fit indices*

	Boys' Models				Girls' Models			
Number of classes	1	2	3	4	1	2	3	4
Log likelihood	−7571.54	−74468.23	**-7408.45**	−7390.15	−7475.07	−7400.91	**-7321.92**	−7404.95
No. free parameters	34	44	**55**	66	34	45	**57**	64
BIC	15360.46	15177.78	**15168.54**	15202.28	15166.30	15087.93	**15006.23**	15216.79
N-Adj BIC	15252.52	15038.09	**14993.93**	14992.75	15058.36	14945.07	**14825.28**	15013.62
AIC	15211.08	14984.46	**14926.89**	14912.31	15018.13	14891.83	**14757.84**	14937.89
Entropy	*a*	0.85	**0.75**	0.82	*a*	0.83	**0.78**	0.63
LRT *p* value for K–1 classes	*a*	.001	**0.028**	*b*	*a*	0.032	**0.002**	0.707

Note: Models with the best fit are shown in bold. BIC, Bayesian information criteria; N-Adj BIC, sample-size adjusted BIC; AIC, Akaike information criteria; LRT *p* value, Lo, Mendell, and Rubin (2001) likelihood ratio test *p* value.
[a]Not estimable for a 1-class model.
[b]Not estimable.

positive quadratic slope (slowing growth over time). To add clinical interpretability to these findings, consider that a boy in Class 3 has a 0% chance of having a preadolescent CBCL internalizing *T* score in the clinical range (*SE* = 0.00), but a boy in Class 1 has a 23% chance (*SE* = 0.09, $p < .01$, OR = infinity).[7]

Girls' internalizing. For girls, the BIC, AIC, log likelihood, and LMR-LRT all found that a two or more class (LGMM) model provided a better fit to the data than the single-class (LGM) model. The next step was to compare two-, three-, and four-class LGMM models. These criteria favored three-classes as the best-fitting model, although comparisons were made with an altered four-class model (that only supported two quadratic slope factors). Although entropy for the three-class model is slightly lower than for the two-class model, Table 3 shows that mean posterior probabilities for each class of the girls' three-class model well exceed adequacy standards. Thus, a three-class model was accepted as the best-fitting model.

Initial internalizing scores for each class of the girls' best-fitting model, shown in Figure 2 and in Table 3, were significantly different from 0. Class 1 (21%, $N = 121$) followed a mainly *elevated-stable* trajectory, although it had a small but significant negative linear slope. Class 2 (10%, $N = 57$) followed a *decreasing/increasing* trajectory, corresponding with a significant negative linear slope and a significant positive quadratic slope. Class 3 included the majority of girls (69%, $N = 399$), and showed a *low* trajectory that decreased somewhat from low to near zero. Clinical interpretability of these classes is aided by considering that a girl in Class 3 has a 0% chance of having an internalizing *T* score in the clinical range (*SE* = 0.00), but a girl in Class 1 has a 12% chance (*SE* = 0.04, $p < .05$, OR = infinity).[7]

In sum, growth mixture results describe latent classes of boys and girls following inter-

7. Low or zero cell counts for children with elevated CDI scores caused some odds ratios to go to zero or infinity, limiting their interest as measures of effect size.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

Table 3. *Growth factor parameter estimates and model-based (posterior) probabilities*

Model	Class	Intercept	Linear Slope	Quadratic Slope	Posterior Probability
Boys' best fitting	1: Elevated–stable	4.38** (0.26)	0.01 (0.09)		0.87
	2: Decrease/increase	3.09** (0.23)	−0.90** (0.11)	0.13** (0.02)	0.86
	3: Low	2.90** (0.11)	−0.76** (0.05)	0.07** (0.01)	0.91
Girls' best fitting	1: Elevated–stable	4.08** (0.22)	−0.13** (0.04)		0.89
	2: Decrease/increase	3.88** (0.32)	−1.50** (0.16)	0.23** (0.02)	0.85
	3: Low	2.78** (0.10)	−0.73** (0.05)	0.07** (0.01)	0.92

Note: Standard errors are in parentheses.
**$p < .01$.

nalizing trajectories of strikingly different shape (stable, increasing, decreasing) and magnitude (high, moderate, low). For both genders, growth mixture models with three classes were statistically preferred over conventional single-class growth models. *If* conventional single-class models had been adopted instead, we would have yielded the following findings. Conventional growth curve models for each gender depict an internalizing trajectory that starts at a level significantly above zero and declines linearly over time. According to this single-class model, girls have a 2% chance of having an elevated CBCL *T* score in preadolescence, and boys a 4% chance.

External validity of latent trajectories of internalizing symptoms

To examine the external validity of extracted latent trajectory classes, we investigated whether class membership predicted preadolescent CDI scores, and/or whether class membership was predicted by mothers' own postpartum symptomology. As hypothesized, girls' and boys' membership in at-risk (*elevated-stable* or *decreasing/increasing*) trajectory classes did predict child-reported depressive symptoms at age 11. Specifically, a girl or boy in the low Class 3 had a 0% chance of being in the CDI clinical range ($SE = 0.00$ and 0.01, respectively), yet a boy in Class 2 had a 8% chance ($SE = 0.03$) and a girl in Class 2 had a 6% chance ($SE = 0.04$) of being in the clinical range ($p < .01$ for both). In addition a girl in Class 1 has a 6% chance of having a clinically elevated CDI score ($SE = 0.03$, $p < .01$). Odds for having high-CDI for boys Class 3 versus 2 is 13.25; odds for having high-CDI for girls Class 2 or 3 versus 1 is infinity.[7]

Although proportions of children in stably elevated classes showing clinically relevant outcomes were small in absolute terms, this does not particularly point to a low true-positive rate. The prior probability of a girl from the studied population evidencing an elevated CDI score is obtained from the general population prevalence rate for depression among prepubertal children (2% for depression, Institute of Medicine, 1989, or 1.3%/0.04%, for depression/dysthymia, respectively; Costello et al., 1988). This prevalence rate resembles the 1.6% rate of having an elevated CDI score in the NICHD ECCYD sample collapsing across classes. Application of Bayes' rule for girls shows that the probability of being classified as elevated stable, if depressed, is .63, whereas the probability of being classified as low stable, if depressed, is 0. Moreover, if class membership were uninformative, we would expect 0.002% of girls in the elevated class to have a CDI elevation, whereas we observed 6% had such.

As hypothesized, children's membership in at-risk trajectory classes was intergenerationally predicted from higher scores on mothers' postpartum symptoms. Allowing maternal covariates to influence growth factors directly and indirectly through the latent classification variable significantly improved model fit compared to the three-class model with distal out-

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

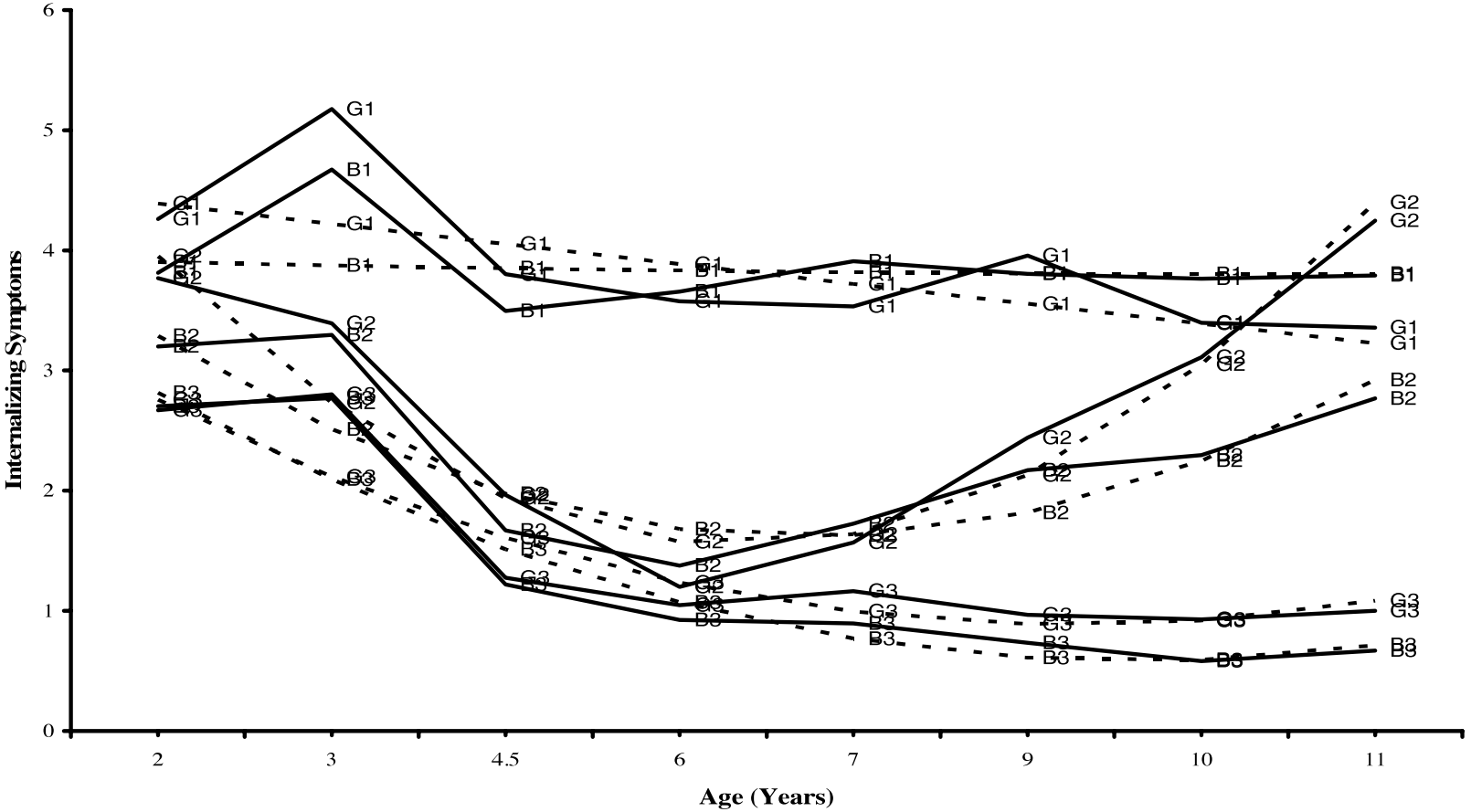


Figure 2. Developmental trajectories of boys' and girls' internalizing symptoms from ages 2 to 11 years. The three boys' trajectory classes are denoted B1, B2, and B3. The three girls' trajectory classes are denoted G1, G2, and G3. Trajectories of sample means are shown with solid lines and corresponding trajectories of model estimated means are shown with dotted lines. Proportions of children in each trajectory class are as follows: B1 (boys' elevated-stable) = 13%, B2 (boys' decreasing/increasing) = 22%, B3 (boys' low) = 65%, G1 (girls' elevated-stable) = 21%, G2 (girls' decreasing/increasing) = 10%, G3 (girls' low) = 69%.

comes, for boys, χ^2 (25, 646) = 496.04, $p <$ .001 and for girls χ^2 (25, 608) =760.14, $p <$ .001. Maternal covariates had direct effects on the intercept and slope growth factors, detailed in the next section. More interestingly, above and beyond these direct effects, maternal variables predicted class membership. Given a mother at mean levels of detachment, nonnegativity, depression, and anxiety, the probability of being in the *elevated-stable* Class 1 is 21% for girls and 12% for boys; the probability of being in the *decreasing/increasing* Class 2 is 9% for girls and 21% for boys, and the probability of being in the *low* Class 3 is 71% for girls and 67% for boys. Yet, for every standard deviation increase in maternal depression, both genders have significantly greater odds of being in Class 1 versus 3 (OR = 1.84, *SE* = 0.34, 95% CI = 1.17–2.51 for girls and OR = 1.57, *SE* = 0.16, 95% CI = 1.25–1.89 for boys) or Class 2 versus 3 (OR = 2.00, *SE* = 0.416, 95% CI = 1.58–2.42 for girls and OR = 1.63, *SE* = 0.25, 95% CI = 1.13–2.13 for boys), holding other maternal covariates constant. These results are suggestive of external validity but not inconsistent with classes being cut points on a continuous distribution rather than true subpopulations.

Gender differences in internalizing symptom trajectories

Previous sections have provided the following descriptive information about gender differences in internalizing symptoms. Girls and boys each evidence three trajectory types, but with differing proportions. More girls than boys follow a elevated-stable trajectory. More boys than girls followed a decreasing/increasing trajectory. Yet, there is a similar breakdown of one-third of children in at risk (*high* or *decreasing/increasing*) versus nonrisk (*low*) trajectories.

Inspection of Figure 2 also suggests gender differences in intercept and slopes, especially for the girls' Class 2 versus the boys' Class 2. This was tested inferentially with a multiple group model formed by combining best-fitting three-class LGMM for each gender. Then the (a) coefficients of intercept and slope factors on maternal covariates, (b) covariance between intercept and slope factors, and (c) within-class residuals were all constrained to be equal across gender. This multiple group model showed good to excellent classification quality (posterior probabilities range = .78–.93, entropy = .83).

First, significant differences were found across gender in class-specific initial values and rates of change in internalizing symptoms. Specifically, nested likelihood ratio tests indicated that boys' class intercepts and slopes were significantly different from their counterpart girls' class intercepts and slopes. Constraining each class's intercept to be equal across gender significantly decreased model fit, χ^2 (1, 1175) = 7.00, $p <$ .01, as did constraining each class's slopes to be equal across gender, χ^2 (4, 1175) = 64.56, $p <$ .001. This was the case even when separate LRTs were performed constraining the slope for one class at a time across gender.

Second, significant differences were not found across gender in class-specific probability of preadolescent outcomes. The probability of being classified in the clinical CDI or clinical CBCL *T*-score range in preadolescence was not significantly different for each class-type across gender. This is shown by the nonsignificant chi-square value upon placing equality constraints on the boy's and girl's thresholds, χ^2 (6, 1175) = 7.61, *ns*.

Third, there were both similarities and differences in the effect of maternal covariates directly on growth factors, for boys and girls. Similarities across gender were that maternal depression predicted higher intercepts for boys, B = 0.03 (0.01), β = .211 $p <$ .05, and girls, B = 0.04 (0.01), β = .202, $p <$ .05. Differences across gender were that maternal separation anxiety predicted higher intercepts, B = 0.02 (0.01) β = .181, $p <$ .05, but lower slopes, B = −0.01 (0.003), β = −0.316, $p <$ .01, only for boys. Maternal lack of negativity predicted lower intercepts, B = −0.27 (0.08), β = −0.135, $p <$ .01, and maternal depression predicted lower slopes, B = −0.012 (0.005), β = −0.315, $p <$ .01, only for girls. These results suggest that postpartum maternal depression for girls and postpartum maternal anxiety for boys heighten children's initial risk of internalizing problems, but these detri-

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

mental effects do not sustain over time. Overall, gender differences were found in class-specific growth intercepts and slopes, and in effects of maternal covariates, but not in trajectory shape or probability of preadolescent outcomes.

Discussion

Traditionally, the course of children's internalizing problems has been tracked from late childhood onward, as if these problems follow a common developmental pathway that varies more so in rate than in qualitative functional form. Not only can this approach obscure the heterogeneity in course that has been long theorized to exist (e.g., Robins, 1966; Rutter, 1989), but it may also lead us to underestimate variation in the onset of gender differences and underestimate long-term relevance of early sociocontextual risks for internalizing pathology. These issues motivated the present investigation of the course of internalizing problems from preschool to preadolescence using methods that more flexibly accommodate heterogeneity in behavioral trajectories.

Our first focus was on describing longitudinal patterns of change. Overall, three-class latent growth mixture models provided best-fitting, substantively compelling, and theoretically consistent depictions of change. Our findings integrated several contradictory findings from prior variable-oriented studies, as follows. Although the majority of children did show stable symptoms from ages 2 to 11 (in support of Keiley et al., 2003), other children did show decreasing symptoms in early childhood (in support of Gazelle & Ladd, 2003) and increasing symptoms in later childhood (in support of Colder et al., 2002). It is likely that the trajectory patterns we observed similarly exist in each of the other three data sets, in different proportions. It seems that portraying change via one common trajectory (and interpreting shifts in slope and intercept specific to ad hoc configurations of predictor values; Bauer & Shanahan, in press) may not as well facilitate integrating findings across studies into a coherent depiction of longitudinal change. We urge researchers to consider more cross-study comparisons of proportions of children probabilistically following (increasing or decreasing or stable) latent trajectory classes, which could help accumulate knowledge of predictors discriminating specific trajectory classes (e.g., Broidy et al., 2003).

Previous studies evidencing stability in internalizing pathology over time using variable-oriented methods have been unable to distinguish whether said stability occurs at all levels of the behavior, and for whom (e.g., Keenan et al., 1998; Mesman & Koot, 2000). Using LGMM we clarify that symptom stability does occur at both the high and low behavioral extremes of the observed range in a community sample, for both boys and girls, from 2 to 11 years. Our hypothesis that children with moderate symptomology would show more instable/modulable patterns of symptom change over time was supported. Their symptoms may vary more so in response to normative environmental stressors and sensitive periods (e.g., the pubertal transition) than do the more-entrenched symptoms from the persisting class (who had a greater accumulation of early risk factors). Our hypothesis that some children would show stable, homotypically internalizing problems from 2 to 11 was supported. This finding downwardly extends evidence that earlier onset of elevated symptomology is associated with a more protracted or severe longitudinal course (e.g., Barlow, 1988; Kovacs et al., 1984). Generalizability of these patterns of relative elevation, stability, and change in internalizing symptoms beyond this normative community context certainly requires replication in clinical samples. Along with Luby (2000) and Angold and colleagues (2005), however, this finding does counter the perception that preschoolers and toddlers are mainly cognitively limited to exhibiting "masked depression," or separation anxiety or externalizing forms of distress.

Our second focus was testing whether maternal internalizing symptoms in the postpartum period could directly predict the dynamic course of child internalizing symptoms from 2 to 11, and, above and beyond these main effects, whether early maternal risk also predicted internalizing trajectory class membership. Direct effects of the set of (mother-

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

reported and observer-reported) maternal symptoms on the slopes and intercepts of (mother-reported) internalizing for girls and boys speak to the longstanding impact of these early risk factors. Yet, substantive interpretation of these results is strongly qualified by the lack of multiple methods of report for both maternal and child psychopathology, as discussed subsequently. Nevertheless, having a postnatally detached, depressed, or anxious mother could render successful adaptation at future milestones less likely by, for example, disrupting the normative organization of cognitive and emotional information at early sensitive periods (e.g., Field et al., 1988; Hsu, 1996). This might longitudinally potentate what has been termed a "depressogenic organizational style" (Cicchetti et al., 2000).

The additional ability of the maternal risk variables (namely maternal depression) to differentially predict internalizing symptom growth via class membership for some classes provided suggestions of external validity of these latent classes. Other indications were found in the classes' ability to differentially predict clinically significant CDI and CBCL scores in preadolescence. The odds ratios comparing likelihood of clinically relevant outcomes for low versus elevated trajectories, however, would have been much more informative if more children had shown clinically significant scores on outcome variables, and if we had used latent outcome measures. The outcome measures should not be thought to "define" the trajectories, which themselves could be specific to internalizing for some individuals but may be "generic pathways that contribute to a range of dysfunctions and disorders" for others (Cicchetti & Toth, 1995, p. 375).

Yet, no maternal health covariates were sensitive enough to discriminate between the *elevated-stable* (Class 1) and the *decreasing/increasing* (Class 2) trajectories. Maternal depression only was found only to distinguish both risk classes from the low class (Class 3). Thus, future research is needed to test whether genetic, child, or family predictors, as well as observer-reported maternal diagnosis, can discriminate between elevated-stable and decreasing/increasing classes. This undertaking will shed light on whether these latent classes should more accurately be viewed as statistical heuristics to aid in classifying and identify individuals at risk, or whether they correspond with truly distinct subpopulations (Bauer & Curran, 2003; Múthen, 2004). Regardless, class membership is helpful in identifying symptom thresholds that correspond with higher probabilities for relevant outcomes.

The final focus of this study was on investigating gender differences in developmental pathways of child internalizing behavior. Prominent gender differences in internalizing pathology have been thought nonexistent until the teen years (Zahn-Waxler, Klimes-Dougan, & Slattery, 2000) without enough methodologically rigorous examination of this hypothesis. By modeling separate dynamic growth trajectories for boys and girls and then inferentially comparing these on multiple fronts, we aimed to elucidate a more nuanced portrayal of the emergence of these gender differences.

First, in contrast to Sommers (2000), there was no evidence of girls' relatively greater socioemotional strengths buffering them from exhibiting early internalizing pathology. The number, shape, and predictive validity of internalizing trajectory classes were similar across gender (i.e., Keenan & Shaw, 1997; Rudolph, 2002), despite sufficient power to detect potential differences (see Eggleston, Laub, & Sampson, 2004). However, significant gender differences were found in the covariates of latent growth factors, consistent with the findings of Colder and colleagues (2002) and Brendgen and colleagues (2001). Specifically, mothers' self-reported postpartum depressive symptoms predicted higher age 2 values of internalizing symptoms across gender, but lack of negativity buffered this effect only for girls and maternal separation anxiety added to this effect only for boys.[8] It is notable that differential effects of maternal symp-

8. Detrimental effects of having a mother with postpartum symptomology significantly diminished over time. This was unanticipated but fits theories suggesting symptoms need to persist to negatively impact the longitudinal sequellae of child internalizing symptoms (e.g., Philipps & O'Hara, 1991).

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

tomology on boys' and girls' internalizing trajectories were detectable in a community sample not selected on maternal psychopathology, and despite the time lag between measurement of mother and child symptoms.

The most interesting point, from the present results, is that the proportion of boys and girls occupying each trajectory class, and each class's initial value and rate of change, varied significantly across gender. Although equal proportions of boys and girls exhibited low behavior, nearly twice as many girls showed *elevated-stable* behavior than boys and nearly twice as many boys showed decreasing/increasing behavior. If future research finds that more girls than boys follow elevated-stable internalizing trajectories in *clinical* samples, relative to diagnostic thresholds, we would want to explore why these girls are often not identified until adolescence. Our result parallels Dekovic and colleagues' (2004) finding that self-reported girls' internalizing trajectories are higher in magnitude than boys', but follow similar growth patterns (albeit among older children aged 13–16). Moreover, our girls' decreasing/increasing class shows lesser decreases and greater increases than their counterpart boys' class. This eventuates in a discrepancy of approximately two symptoms by age 11, as anticipated by extant developmental theory (e.g., Costello et al., 1988). It is important for future research to identify peer or familial risk factors that differentiate those girls who show relatively elevated-stable internalizing symptoms from those whose elevated symptoms are delimited to specific developmental phases: toddlerhood and then preadolescence (i.e., Class 2, *decreasers/increasers*).

Significant findings should be considered in the context of a number of study limitations. First, should longitudinal diagnostic information have been available in the NICHD data set, it would have strengthened findings to specify trajectory classes for anxiety and depressive symptoms separately. This study serves as an initial attempt to chart continuity of broadband internalizing symptoms and uses a combination of anxiety and depressive symptoms to do so, as there is evidence that these are not fully differentiated in early childhood (Nottelmann & Jensen, 1995).

Second, the sole reliance on maternal-reported CBCL scores to measure child internalizing problems potentially elevated our estimates of the stability of child problems, if consistent reporting biases contributed to stability of latent growth parameters. Part of the association of maternal ratings of maternal depression and maternal ratings of child internalizing could be because of depressed mothers overstating and overgeneralizing both their children's and/or their own problems (Chilcoat & Breslau, 1996; cf. Biederman, Mick, & Faraone, 1998). Maternal reporting biases may also have accounted for some of the gender differences observed, if mothers differentially perceive internalizing problems in boys versus girls (see Murray & Cooper, 1997). Replication of the gender differences found here in a study utilizing multiple reporters and multiple methods of reporting for each gender would increase our confidence in the substantive meaning of these findings. This study would also have benefited from linkage of maternal-reported child internalizing trajectories to child-reported and teacher-reported symptoms (following Keiley et al., 2003) as a further check on external validity. Because children's self-reported internalizing symptoms are only valid from ages 5 to 6 on (Ialongo, Edelsohn, & Kellam, 2001), they could not have been used as the primary outcome variable (required to be invariant across ages 2–11). If such self-report measures had been available, they could have importantly served as time-varying covariates of maternal reported symptoms from ages 5 to 11.

Third, this study established measurement invariance of the repeated, dependent variable both theoretically (Achenbach 1991, 1992) and empirically, using Cronbach's alpha. This has been the standard approach to date for growth modeling studies (e.g., Gilliom & Shaw, 2004) in early childhood. More rigorous testing of this assumption would have entailed estimating a longitudinal confirmatory factor analysis model, and then linking this to our growth model. It is unclear, however, whether we clinically conceptualize internalizing symptoms as having as fastidiously consistent a presentation over time as would be operationalized by constraining item loadings and residual vari-

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

ances to be equal across time (see Cicchetti & Toth, 1995, p. 377). The one study we found that attempted to demonstrate this was only partially successful (Colder et al., 2002). Heterotypic presentation of the internalizing construct over time interferes with growth modeling assumptions, but may need to be elucidated and wrestled with as studies of internalizing developmental pathways proliferate. Last, as keenly highlighted by Howe (2004), the year or more spacing of assessment points may have inflated the appearance of systematic, unidirectional patterns of change. Repeated examination of symptoms on a narrower time scale, especially around transition periods, could have revealed more erratic or discontinuous patterns of change for girls and boys. This is an important area for future investigations.

Conclusion

In conclusion, by allowing heterogeneity in our model of the developmental course of internalizing problems (Nagin, 1999; Raudenbush, 2001), we were able to integrate some discrepant findings in prior work, and resolve an ambiguity about the nature of stability in the internalizing construct. By including preschoolers in these analyses (following Angold et al., 2005, and Luby et al., 2003), we were able to examine early-starter models of internalizing pathology and shed new light on debates about the onset of gender differences in internalizing pathology. We highlighted the relevance of maternal-reported maternal postpartum depression for the longitudinal sequellae of maternal-reported internalizing symptoms. Because maternal postpartum symptoms only modestly differentiated internalizing pathways, more work is needed to decipher whether latent classes of children identified using latent growth mixture methods do correspond to qualitatively different subpopulations. Either way, they appear promising both for investigating the longitudinal course of internalizing problems and for facilitating comparison and integration of results across studies.

References

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

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Albano, A., Chorpita, B., & Barlow, D. (1996). Childhood anxiety disorders. In E. Mash & R. Barkley (Eds.), *Child psychopathology*. New York: Guilford Press.

Angold, A., Egger, H., Erkanli, A., & Keeler, G. (2005). *Prevalence and comorbidity of psychiatric disorders in preschoolers attending a large pediatric service*. Unpublished manuscript.

Barlow, D. (1988). *Anxiety and its disorders: The nature and treatment of anxiety and panic*. New York: Guilford Press.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363.

Bauer, D., & Shanahan, M. (in press). Modeling complex interactions: Person-centered and variable-centered approaches. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development*. Mahwah, NJ: Erlbaum.

Beitchman, J., Wekerle, C., & Hood, J. (1987). Diagnostic continuity from preschool to middle childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *26*, 694–699.

Biederman, J., Mick, E., & Faraone, S. (1998). Biased maternal reporting of child psychopathology? *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*, 10–11.

Birmaher, B., Williamson, D., Dahl, R., Axelson, D., Kaufman, J., & Dorn, L., et al. (2004). Clinical presentation and course of depression in youth: Does onset in childhood differ from onset in adolescence? *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*, 63–70.

Bradley, R. H. (1994). The Home Inventory: Review and reflections. In H. W. Reese (Ed.), *Advances in child behavior and development* (Vol. 25, pp. 241–288). Orlando, FL: Academic Press.

Brendgen, M., Vitaro, F., Bukowski, W., Doyle, A., & Markiewicz, D. (2001). Developmental profiles of peer social preference over the course of elementary school: Associations with trajectories of externalizing and internalizing behavior. *Developmental Psychology*, *37*, 308–320.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, *39*, 222–245.

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.

Campbell, S., Cohn, J., Flanagan, C., Popper, S., & Meyers, T. (1992). Course and correlates of postpartum

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

depression during transition to parenthood. *Development and Psychopathology*, *4*, 29–47.
Cantwell, D., & Baker, L. (1989). Stability and natural history of *DSM-III* childhood diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, *29*, 691–700.
Carlson, G., & Kashani, J. (1988). Phenomenology of major depression from childhood through adulthood: Analysis of three studies. *American Journal of Psychiatry*, *145*, 1222–1225.
Chilcoat, H., & Breslau, N. (1996). Does psychiatric history bias mothers' reports? An application of a new analytic approach. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 971–979.
Cicchetti, D., & Rogosch, F. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, *8*, 597–600.
Cicchetti, D., Rogosch, F., & Toth, S. (2000). The efficacy of toddler–parent psychotherapy for fostering cognitive development in offspring of depressed mothers. *Journal of Abnormal Child Psychology*, *28*, 135–148.
Cicchetti, D., & Toth, S. (1995). Developmental psychopathology and disorders of affect. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology. Volume I: Theory and Methods* (pp. 369–420). New York: Wiley.
Cohn, J., Campbell, S., Matias, R., & Hopkins, J. (1990). Face to fact interactions of postpartum depressed and nondepressed mother–infant pairs. *Developmental Psychopathology*, *26*, 15–23.
Colder, C., Mott, J., & Berman, A. (2002). The interactive effects of infant activity level and fear on growth trajectories of early childhood behavior problems. *Development and Psychopathology*, *14*, 1–23.
Costello, E., Costello, A., Edelbrock, C., Burns, B., Dulcan, M. K., Brent, D., et al. (1988). Psychiatric disorders in pediatric primary care. *Archives of General Psychiatry*, *45*, 1107–1116.
Dekovic, M., Buist, K., & Reitz, E. (2004). Stability and changes in problem behavior during adolescence: Latent growth analysis. *Journal of Youth and Adolescence*, *33*, 1–12.
Egeland, B., & Sroufe, L. (1981). Developmental sequelae of maltreatment in infancy. *New Directions for Child Development*, *11*, 77–92.
Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, *20*, 1–26.
Field, T., Healy, B., Goldstein, S., Perry, S., Bendell, D., Schanberg, S., et al. (1988). Infants of depressed mothers show "depressed" behavior even with nondepressed adults. *Child Development*, *59*, 1569–1579.
Field, T., Lang, C., Martinez, A., Yando, R., Pickens, J., & Bendell, D. (1996). Preschool follow-up of infants of dysphoric mothers. *Journal of Clinical Child Psychology*, *25*, 272–279.
Galambos, N., Barker, E., & Almeida, D. (2003). Parents *do* matter: Trajectories of change in externalizing and internalizing problems in early adolescence. *Child Development*, *74*, 578–594.
Gazelle, H., & Ladd, G. (2003). Anxious solitude and peer exclusion: A diathesis-stress model of internalizing trajectories in childhood. *Child Development*, *74*, 257–278.
Gilliom, M., & Shaw, D. (2004). Co-development of externalizing and internalizing problems in early childhood. *Development and Psychopathology*, *16*, 313–333.
Hock, E., DeMeis, D., & McBride, S. (1987). Maternal separation anxiety: Its role in the balance of employment and motherhood in mothers of infants. In A. Gottfried & A. Gottfried (Eds.), *Maternal employment and children's development: Longitudinal research* (pp. 191–229). New York: Plenum Press.
Hock, E., Gnezda, M., & McBride, S. (1983). *The measurement of maternal separation anxiety*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Detroit.
Hollenstein, T., Granic, I., Stoolmiller, M., & Snyder, J. (2004). Rigidity in parent–child interactions and the development of externalizing and internalizing behavior in early childhood. *Journal of Abnormal Child Psychology*, *32*, 595–607.
Howe, G. (2004). Studying the dynamics of problem behavior across multiple time scales: Prospects and challenges. *Journal of Abnormal Child Psychology*, *32*, 673–679.
Hox, J., & Maas, C. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, *8*, 157–174.
Hsu, H. (1996). Antecedents and consequences of separation anxiety in first time mothers: Infant, mother, and socio-contextual characteristics. *Infant Behavior and Development*, *27*, 113–133.
Ialongo, N., Edelsohn, G., & Kellam, S. (2001). A further look at the prognostic power of young children's reports of depressed mood and feelings. *Child Development*, *72*, 736–747.
Institute of Medicine. (1989). *Research on children and adolescents with mental, behavioral, and developmental disorders*. Washington, DC: National Academy Press.
Jensen, P. S., & Watanabe, H. (1999). Sherlock Holmes and child psychopathology assessment approaches: The case of the false-positive. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*, 138–147.
Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behavior. *Psychological Bulletin*, *121*, 95–113.
Keenan, K., Shaw, D., Delliquadri, E., Giovannelli, J., & Walsh, B. (1998). Evidence for the continuity of early problem behaviors: Application of a developmental model. *Journal of Abnormal Child Psychology*, *26*, 441–450.
Keenan, K., Shaw, D., Walsh, B., Delliquadri, E., & Giovannelli, J. (1997). *DSM-III-R* disorders in preschool children from low-income families. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 620–627.
Keiley, M., Lofthouse, N., Bates, J., Dodge, K., & Pettit, G. (2003). Differential risks of covarying and pure components in mother and teacher reports of externalizing and internalizing behavior across ages 5 to 14. *Journal of Abnormal Child Psychology*, *31*, 267–284.
Kovacs, M. (1992). *Children's Depression Inventory–Short Form*. North Tonawanda, NY: Multi-Health Systems.
Kovacs, M., Feinberg, T., Crouse-Novak, M., Paulauskas, S., Pollock, M., & Finkelstain, R. (1984). Depressive disorders in childhood II: A longitudinal study of the risk for a subsequent major depression. *Archives of General Psychiatry*, *41*, 643–649.
Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture, *Biometrica*, *88*, 767–778.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press

Luby, J. L. (2000). Depression. In C. H. Zeanah (Ed.), *Handbook of infant mental health* (2nd ed., pp. 382–396). New York: Guilford Press.

Luby, J. L., Heffelfinger, A. K., Mrakotsky, C., Brown, K. M., Hessler, M. J., & Wallis, J. M. (2003). The clinical picture of depression in preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*, 340–348.

MacCallum, R., Zhang, S., Preacher, K., & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40.

Matthey, S., & Petrovski, P. (2002). The children's depression inventory: Error in cutoff scores for screening purposes. *Psychological Methods*, *14*, 146–149.

McLennan, J., & Offord, D. (2002). Should postpartum depression be targeted to improve child mental health? *Journal of the American Academy of Child & Adolescent Psychiatry*, *28*.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543.

Mesman, J., & Koot, H. (2000). Child-reported depression and anxiety in preadolescence: II. Preschool predictors. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 1379–1386.

Mitchell, J., McCauley, E., Burke, P, & Moss, S. (1988). Phenomenology of depression in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *27*, 12–20.

Muris, P., Merckelbach, H., Gadet, B., & Moulaert, V. (2000). Fears, worries, and scary dreams in 4- to 12-year-old children: Their content, developmental pattern, and origins. *Journal of Clinical Child Psychology*, *29*, 43–52.

Murray, L., & Cooper, P. (1997). Postpartum depression and child development. *Psychological Methods*, *27*, 253–260.

Múthen, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). New York: Sage.

Múthen, B., & Asparouhov, T. (2002). *Using Mplus Monte Carlo simulations in practice: A note on non-normal missing data in latent variable models. Web notes no. 2*. Los Angeles: Múthen & Múthen.

Múthen, L. K., & Múthen, B. O. (1998–2004). *Mplus user's guide* (3rd ed.). Los Angeles: Múthen & Múthen.

Múthen, L. K., & Múthen, B. O. (2002). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.

Múthen, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

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

NICHD Early Child Care Research Network. (2004a). Affect dysregulation in the mother–child relationship in the toddler years: Antecedents and consequences. *Development and Psychopathology*, *16*, 43–68.

NICHD Early Child Care Research Network. (2004b). Trajectories of physical aggression from toddlerhood to middle childhood. *Monographs of the Society for Research in Child Development*, *69*.

Nottelmann, E. D., & Jensen, P. S. (1995). Comorbidity of disorders in children and adolescents: Developmental perspectives. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 17, pp. 109–155). New York: Plenum Press.

Philipps, L., & O'Hara, M. (1991). Prospective study of postpartum depression: 4 1/2 year follow-up of women and children. *Journal of Abnormal Psychology*, *100*, 151–155.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*, 501–525.

Robins, L. (1966). *Deviant children grown up*. Baltimore, MD: Williams & Williams.

Rudolph, K. (2002). Gender differences in emotional responses to interpersonal stress during adolescence. *Journal of Adolescent Health*, *30*, 3–13.

Rutter, M. (1989). Pathways from childhood to adult life. *Journal of Child Psychology and Psychiatry*, *30*, 23–51.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford Press.

Sommers, C. H. (2000). *The war against boys: How misguided feminism is hurting our young men*. New York: Simon & Schuster.

von Eye, A., & Bergman, L. R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology*, *15*, 553–580.

Zahn-Waxler, C., Klimes-Dougan, B., & Slattery, M. (2000). Internalizing problems of childhood and adolescence: Prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology*, *12*, 443–466.

https://doi.org/10.1017/S0954579407070174 Published online by Cambridge University Press