

A Daily Diary Study on Adolescent Emotional Experiences: Measurement Invariance and Developmental Trajectories

Dominique F. Maciejewski and Pol A. C. van Lier
Vrije Universiteit Amsterdam and EMGO Institute for Health
and Care Research, Amsterdam, The Netherlands

Susan J. T. Branje
Utrecht University

Wim H. J. Meeus
Utrecht University and Tilburg University

Hans M. Koot
Vrije Universiteit Amsterdam and EMGO Institute for Health
and Care Research

Adolescence is an important time for emotional development. Recently, daily diary methods are increasingly employed in research on emotional development and are used to explore the development of and sex differences in emotions during adolescence. However, before drawing conclusions about sex differences and developmental trends, one needs to ensure that the same construct is measured across sex and time. The present study tested measurement invariance of daily emotion assessments across sex, short-term (days within weeks) and long-term periods (days across years) in a sample of 394 adolescents (55.6% male) that were followed from ages 13 to 18. Moreover, the study examined the developmental trajectories of adolescent emotional experiences. Adolescents rated their daily emotions (happiness, anger, sadness, anxiety) during each day of a normal school week (Monday to Friday) for 3 weeks per year for 5 years (i.e., 15 weeks \times 5 days = 75 assessments in total). Measurement invariance analyses suggest that the measurement of adolescent daily mood was invariant between boys and girls and across shorter and longer time intervals. Moreover, latent growth curve analyses showed that happiness decreased from early to middle adolescence, whereas anger, sadness, and anxiety increased. Anger returned to baseline toward late adolescence. In contrast, the decrease of happiness and the increase of anxiety leveled off without reversing, whereas sadness continued to increase. The discussion highlights the implications of measurement invariance in research on individual and developmental differences and discusses the findings in light of normative emotional development.

Keywords: emotions, daily diaries, measurement invariance, development, adolescence

Supplemental materials: <http://dx.doi.org/10.1037/pas0000312.supp>

Emotions play an important role in our lives: They notify us of changes in our environment and stimulate actions to cope with these changes to maintain our well-being (Kuppens, 2015). Adolescence is a time in which emotional experiences are subject to significant changes. This may not come as a surprise, considering that adolescence is a period of great transition (Arnett, 1999). Due to the dynamic nature of emotions (Kuppens, 2015), intensive repeated measurements (e.g., daily diaries) have gained popularity in the field of emotion research. These methods have high ecological validity,

minimize recall bias, and provide more reliable estimates than conventional methods that require individuals to summarize their emotions across many situations (Shiffman, Stone, & Hufford, 2008). Despite this, empirical evidence on the development of daily emotions across adolescence using intensive measurements and longitudinal designs is scarce, although studying normative emotion trajectories may have important implications for understanding both normal and abnormal emotional development. Moreover, before drawing conclusions about sex differences or developmental trends, one needs to be

Dominique F. Maciejewski and Pol A. C. van Lier, Department of Clinical, Neuro- & Developmental Psychology, Section Clinical Developmental Psychology, Vrije Universiteit Amsterdam, and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands; Susan J. T. Branje, Research Centre Adolescent Development, Utrecht University; Wim H. J. Meeus, Research Centre Adolescent Development, Utrecht University, and Department of Developmental Psychology, Tilburg University; Hans M. Koot, Department of Clinical, Neuro- & Developmental Psychology, Section Clinical Developmental Psychology, Vrije Universiteit Amsterdam and EMGO Institute for Health and Care Research.

Data of the RADAR study were used. RADAR has been financially supported by main grants from the Netherlands Organisation for Scientific Research (GB-MAGW 480-03-005, GB-MAGW 480-08-006), the Stichting Achmea Slachtoffer en Samenleving (SASS), a grant from the Netherlands Organisation for Scientific Research to the Consortium Individual Development (CID; 024.001.003), and various other grants from the Netherlands Organisation for Scientific Research, the VU University Amsterdam, and Utrecht University. We want to thank Delia Burke for critically reviewing an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Dominique F. Maciejewski, Vrije Universiteit Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. E-mail: d.f.maciejewski@vu.nl

sure that the same construct is measured across sex and time (i.e., one needs to ensure that a construct is *measurement invariant*). If that is not the case, then any interpretation concerning sex differences or developmental changes would be obscured by differences in measurement. To address the short-comings of previous studies, the present study had two goals: First, using longitudinal data from adolescents that were followed from ages 13 to 18, we tested whether adolescent daily emotions (happiness, sadness, anger, and anxiety) were measurement invariant across sex and time. Second, we examined the developmental changes of daily emotions across adolescence, while taking into account possible sex differences.

Developmental Changes in Emotional Experiences Across Adolescence

Adolescence is a period with many developmental challenges. For instance, adolescents have to accommodate to biological changes, develop a stable identity, and deal with transformations in interpersonal relationships (for a review, see Holmbeck, Friedman, Abad, & Jandasek, 2006). The experience of such developmental changes may be tied to the experience of more negative moods during adolescence, such as anxiety (see Weems, 2008). Moreover, advances in cognitive capacities enable adolescents to “see beneath the surface of situations and envision hidden and more long-lasting threats to their wellbeing” (Larson, Richards, & Perry-Jenkins, 1994, p. 86). Indeed, studies indicate that adolescents do not only experience more potentially stressful life events than children, but also more negative emotions in response to such events (Larson & Ham, 1993).

In the past years, research on adolescents’ daily emotional experiences has grown and some studies have explicitly tested developmental changes in emotions during adolescence using daily diary methods. For instance, two cross-sequential studies, in which adolescents provided reports on their daily emotions, indicated that the average emotional state became less positive across adolescence. The decline in positive emotions was steepest in early adolescence and then leveled off toward late adolescence without reversing (Larson, Moneta, Richards, & Wilson, 2002; Moneta, Schneider, & Csikszentmihalyi, 2001). Both studies, however, used bipolar items (e.g., scales ranging from sad to happy) and could therefore not differentiate between different emotions. This is, however, potentially important, because research indicates that positive and negative emotions are relatively independent constructs (Lonigan, Hooe, David, & Kistner, 1999). A more recent diary study that differentiated between positive and negative global affect (but not between specific emotions) showed that only positive affect was significantly lower in older compared to younger adolescents, whereas no difference was found for negative affect (Weinstein, Mermelstein, Hankin, Hedeker, & Flay, 2007). To our knowledge, no study has examined distinct emotions using daily diary methods. However, previous studies using conventional self-report methodologies found significant changes in negative affect across adolescence. Longitudinal studies on the development of aggressive symptoms, an outcome of angry emotions, have shown that such symptoms peak during middle adolescence (Karriker-Jaffe, Foshee, Ennett, & Suchindran, 2008; Meeus, Van de Schoot, Hawk, Hale, & Branje, 2016). In contrast to that, depressed mood has been found to linearly increase across adolescence (Van Oort, Greaves-Lord, Verhulst, Ormel, & Huizink,

2009), especially after age 13, and especially for girls (Cole et al., 2002; Ge, Lorenz, Conger, Elder, & Simons, 1994). Longitudinal studies on anxiety symptoms have shown that generalized anxiety symptoms first decrease in early adolescence, and then increase again from middle adolescence onward (Nelemans et al., 2014; Van Oort et al., 2009).

Apart from developmental changes in different emotions, there may be important sex differences in the normative development of adolescents’ emotions. For instance, theories on sex differences in depressive symptoms hypothesize that girls show greater increases in depressive emotions in adolescence than boys, because they experience more negative life events and have a higher cognitive vulnerability than boys (e.g., girls ruminate more; Hankin & Abramson, 2001). Thus, it is possible that girls experience more dramatic increases in negative moods compared to boys when entering adolescence. Daily diary studies that have examined sex differences in emotional experiences are however not very consistent. Some have found that girls experience more negative emotions (Silk, Steinberg, & Morris, 2003), others that girls experience more positive emotions (Larson et al., 2002), and yet others have found no sex differences at all (Weinstein et al., 2007). Moreover, studies on sex differences in developmental trajectories indicate that boys and girls show similar developmental changes in average affect across adolescence (Larson et al., 2002; Moneta et al., 2001). However, as mentioned, both studies did not differentiate between different emotions.

To address the short-coming of previous studies, which are predominately based on conventional self-reports, have employed cross-sequential or cross-sectional designs, and/or have not examined different emotions, the first aim of the present study was to study sex-specific developmental trajectories of four primary emotions, namely happiness, anger, sadness, and anxiety. The present study focused on primary emotions rather than secondary emotions (e.g., shame, guilt). Primary emotions have some important features that secondary emotions do not possess, such as evolutionary importance (i.e., primary emotions stimulate actions for survival purposes), cross-cultural universality, and ontogenetic primacy (i.e., they develop earlier in life compared to secondary emotions). In contrast, secondary emotions are consequences of social constructions and are dependent on primary emotions (e.g., shame as a socialized response to anger; for more discussion, see Kemper, 1987). Due to the complexity of secondary emotions and their dependence on the social context, the following study focused exclusively on primary emotions.

The Need for Establishing Measurement Invariance

An important prerequisite for making comparisons across sex and time is that the measurement of a construct is equivalent across sex and time. Importantly, the fact that the measurement of a construct, in this case emotions, does not differ across sex or time (i.e., is invariant) does not imply that the construct itself does not differ across sex or time. It implies that demonstrated differences in the construct can be attributed to true differences, and not to differences in measurement. This issue, although often neglected in research on individual differences and development, is addressed by establishing *measurement invariance* (Chen, Sousa, & West, 2005; Gregorich, 2006; Little, Preacher, Selig, & Card, 2007; Vandenberg & Lance, 2000; Widaman, Ferrer, & Conger, 2010). If we cannot be certain that we are

measuring the same construct in the same metric, conclusions regarding individual or developmental differences are flawed (Widaman et al., 2010).

The issue of measurement invariance becomes more complicated when the measurement of a construct is very time-intensive, as it is the case for daily diary methods. Here, changes in the measurement can happen within smaller (i.e., across days) and larger time frames (i.e., across years). For instance, if adolescents have to rate their daily mood across a week, it might be that they perceive items about emotions differently on different days, or that they subjectively redefine their emotional experiences based on their earlier answers. This would mean that the same underlying emotion is no longer being measured and that comparisons or aggregation of emotion scores across days may not be justified.

The same holds for measurement invariance across longer time periods. For instance, before studying developmental differences between early and late adolescence, one needs to ensure that the measurement of emotions is invariant across years. Possible differences between ages may, otherwise, merely represent variability in the measurement over time rather than true developmental changes (Widaman et al., 2010). One could for instance argue that cognitive and emotional processing abilities are still developing in adolescence (Yurgelun-Todd, 2007). Indeed, research has shown that the accuracy and speed of emotion identification and differentiation continuously increases from childhood to young adulthood (Gur et al., 2012). As a result, it is possible that, compared to older adolescents, younger adolescents have more difficulties in correctly identifying their emotions and distinguishing between different emotions, particularly if emotions are more alike (e.g., anxiety and sadness). For instance, research indicates that depressive and anxiety symptoms become more differentiated from each other in middle adolescence (Sterba et al., 2010). Such differences may consequently influence emotion ratings, affect the measurement of emotions across different developmental periods, and lead to biased conclusions about developmental changes.

Similarly, measurement invariance needs to be established across sex to be able to make meaningful comparisons of emotional experiences between boys and girls. If boys and girls interpret their internal emotional states differently, then comparing them on exactly these emotional experiences would have little value. Indeed, it is possible that this is the case. For instance, research indicates that girls attend more to their emotions than boys (Neumann, van Lier, Gratz, & Koot, 2010) and that they perform better on emotion identification and differentiation tasks (Gur et al., 2012), which may also make them better reporters of their internal emotional states. Girls may better understand when a certain emotion is high or low, whereas in boys, emotions may need to cross a certain threshold to be noticed by them. There may also be differences with regard to specific emotions. For instance, boys may have more difficulties reporting on their sadness, because studies have for instance found that they respond in a more hostile and aggressive way to depressed mood compared to girls (Gjerde, 1995). The opposite might be true for anger reports by girls, because girls have been found to suppress their angry emotions to a greater extent than boys (Cox, Stabb, & Hulgus, 2000). Thus, if there are certain emotions that either boys or girls are less comfortable in expressing, then this might lead to differences in the interpretation of these emotions and subsequent differences in the ratings and interrelations of these different emotions. In that

case, possible sex differences in emotions would just be a measurement artifact (Chen et al., 2005; Gregorich, 2006).

In the following, we want to give a short overview about the theory and concept of measurement invariance. Measurement invariance involves hierarchical testing of different measurement invariance forms (for more information of invariance testing, see Chen et al., 2005; Gregorich, 2006; Vandenberg & Lance, 2000). The theoretical assumptions that determine this hierarchical testing are based on the classical test theory. According to this, the response to an item is defined as a linear function of a regression slope (or factor loading), an item intercept, and an error component (or residual variance). In measurement invariance testing, each of the components of the equation is hierarchically constrained to be equal across groups and/or time, to test whether the measurement models are equivalent (Vandenberg & Lance, 2000).

The first step of measurement invariance is to test for *configural invariance*, which tests whether the factor structure is the same across groups and/or time. If configural invariance is established, then this implies that the same items load on the same total score, between groups and/or across time. The second step is to test for *metric invariance* (i.e., additional equal factor loadings). If metric invariance is established, then this implies that items contribute in the same way to the total scores, between groups and/or across time. This makes it possible to compare relationships with other external variables between groups or time-points¹ (Chen et al., 2005; Gregorich, 2006). The third step is to test for *scalar invariance* (i.e., additional equal item intercepts). If scalar invariance is established, then this implies that the starting points in rating emotions are equal. This makes it possible to compare total or latent scores between groups or different time-points, which is a necessary precondition for mean comparisons. The fourth step is to test for *strict invariance* (i.e., additional equal residual variances). If strict invariance is established, then differences in items are fully explained by differences in latent factors/total scores, and not by differences in random or unknown influences. Strict invariance is often difficult to establish, partly because it is rather unlikely that the amount of random influences is equal between groups or across time (Gregorich, 2006; Little et al., 2007; Widaman et al., 2010). However, strict invariance has limited additional value for most research questions (Gregorich, 2006; Little et al., 2007), and scalar invariance is sufficient for most popular longitudinal data analyses (e.g., growth curve models; Khoo, West, Wu, & Kwok, 2006).

Empirical evidence on measurement invariance of daily diary emotion data is virtually nonexistent. There is some research suggesting that the factor structure and factor loadings of positive and negative emotions using the Positive and Negative Affect Schedule do not differ between boys and girls and between children (ages 9–11) and adolescents (ages 12–17; Lonigan et al., 1999). However, to our knowledge, no study to date has examined measurement invariance of daily mood data between boys and girls, days within weeks, and days across years.

¹ Provided metric invariance has also been established for this construct.

The Present Study

The present study had two aims: First, we tested whether adolescent daily emotions were measurement invariant across sex and time. Specifically, we tested for invariance across sex, days within a normal school week (i.e., Monday to Friday), and days across years, using daily emotion data that were collected during a 5-year period (ages 13 to 18). Second, if we indeed found evidence for configural, metric, and scalar invariance across sex and time, we sought to examine developmental changes of emotions and sex differences in these changes across the period of adolescence. Due to the limited research on measurement invariance of daily mood data, we did not form specific predictions about the level of invariance that would be achieved. Based on previous longitudinal studies on developmental trajectories of aggression, depressed mood, and anxiety, we hypothesized initial increases in anger, followed by decreases (i.e., peaks in middle adolescence), increases for sadness, and initial decreases for anxiety in early adolescence, followed by increases toward late adolescence. Because, to our knowledge, no longitudinal study has been previously conducted on happiness, we did not form specific hypotheses about the development of happiness.

We modeled our data using second-order factors and tested for invariance for both specific emotions (i.e., items loading on associated emotions), as well as general negative moods (i.e., emotions loading on general negative mood). This approach was taken because it has been suggested that discrete moods constitute rather general mood factors (e.g., general positive and negative moods; Lonigan et al., 1999), and because several researchers have aggregated repeated mood data from different emotions to construct general mood scores (Weinstein et al., 2007).

Method

Sample and Procedure

For the present study, we used data from the Research on Adolescent Development and Relationships-Young Cohort (RADAR-Y) study, which is an ongoing longitudinal research project in the Netherlands in which adolescents have been followed from age 13 onward. The present study used data from ages 13 to 18 years. The RADAR study has a focus on delinquency development. For this reason, adolescents with borderline scores on teacher-reported externalizing behavior problems at age 12 were oversampled (T score > 60 on the Teacher's Report Form; TRF; Achenbach, 1991). The final sample consisted of 497 adolescents, of which 291 adolescents were at average risk (TRF/6–18 years scores on externalizing behavior below the borderline ranges; 165 boys and 126 girls) and 206 adolescents were at increased risk for externalizing behavior (TRF/6–18 borderline or higher scores; 118 boys and 88 girls). The medical ethical committee of the University Medical Centre Utrecht approved the study and all participants provided written informed consent (more information on the sample selection can be found in Neumann, van Lier, Frijns, Meeus, & Koot, 2011).

During the course of 5 years, 486 of these 497 adolescents participated in 15 Internet assessment weeks, which lasted 5 days each (Monday to Friday; 3 Internet assessment weeks within each year). Thus, in total there were 75 Internet assessments across the

5 years. Data were collected in June (Month 6), September (Month 9), and December (Month 12) from 2006 until 2010. The RADAR project also included annual home visits, which took place mostly around February and March (Months 2 and 3), but which were not part of the present study. The measurement months were intentionally planned in that order, so that Internet and home assessments would always be separated by approximately three months. Adolescents' mean age at the first Internet assessment week was 13.31 years ($SD = 0.45$). Each Internet assessment covered the weekdays of one normal school week (Monday to Friday). In these Internet assessments, adolescents were asked to rate their daily happiness, anger, anxiety, and sadness for that particular day. To complete the Internet assessments, adolescents logged on to the RADAR website. Adolescents were reminded of participation by e-mail invitations, which were sent to them at approximately 5:30 p.m. In the case that adolescents had not completed the assessments 1.5 hr after the first e-mail invitations, they received reminder e-mails. After an additional 1.5 hr, adolescents received text messages and phone calls as additional reminders. Adolescents received approximately €10 (equivalent to US\$11) for participation in each Internet assessment wave.

Missing Data

In the present study, 79.8% of adolescents were still participating in the last year of the Internet assessments. Adolescents that dropped out of the study in the last assessment year were more likely to be from a low socioeconomic status (SES), $\chi^2(1) = 18.62$, $p < .001$, and to have borderline scores on teacher-reported externalizing problems at age 12, $\chi^2(1) = 6.74$, $p = .009$, compared to adolescents that were still participating in the last assessment year. The two groups however did not differ on sex, $\chi^2(1) = 1.08$, $p = .30$, and emotion levels (happiness, anger, sadness, and anxiety) averaged across the study period, all $ps > .17$. Across all participants and assessment days (i.e., 486 adolescents \times 75 assessment days = 36,450 possible observations), there were 24,064 valid observations, meaning that there were 34.0% missing data observations in total. The average amount of valid assessment days was 49.51 assessment days ($SD = 19.23$) per adolescent.

To minimize the possible influence of missing data on the measurement invariance analyses, we only included adolescents in the analyses that participated in at least two thirds of the Internet assessment weeks (i.e., a minimum of 10 out of 15) across the 5 years.² Of the initial 486 adolescents, 394 were included by virtue of this criterion. Adolescents that were excluded were more likely to be from a low SES, $\chi^2(1) = 9.29$, $p = .002$, and to have borderline scores on teacher-reported externalizing problems at age 12, $\chi^2(1) = 10.45$, $p = .001$. The excluded and final sample did not differ on sex, $\chi^2(1) = 0.85$, $p = .36$, and emotion levels, all $ps > .11$. The final sample consisted of 394 adolescents (55.6% male; 37.1% at risk for externalizing behavior at age 12), who had

² We have replicated part of the measurement invariance analyses reported in the result section with two different inclusion criteria; First, with a more lenient criterion by including all adolescents that participated. Second, with an even stricter criterion by including only adolescents that participated for a minimum of 2 weeks in each of the 5 consecutive years (see Online Resource 4 in the Supplemental Materials).

a predominately middle or high SES (92.0%) and a Dutch-Caucasian background (97.5%).

Adolescent Daily Mood Measure

Adolescents were administered the Daily Mood Device, an Internet version of the Electronic Mood Device (Hoeksma et al., 2000). Using online questionnaires, adolescents rated the intensity of their daily emotional state for happiness, anger, anxiety, and sadness ("Today I feel . . .") on five consecutive school days (i.e., Monday to Friday). Each of the four emotion states was measured using three items (see Table 2 for the items). Items were rated on a 9-point Likert scale ranging from 1 (*not glad/sad/anxious*) to 9 (*glad/sad/anxious*). We calculated Cronbach's alpha for each of the four mood scales for each of the 75 days. Across the study period, Cronbach's α s ranged from .86 to .96 for happiness, from .88 to .96 for anger, from .73 to .94 for anxiety, and from .91 to .98 for sadness, per day.

To check whether adolescents had read the questionnaires properly and did not, for instance, rate positive and negative items all high or low, we calculated how often adolescents rated all 1s or 9s on all 12 items (i.e., three positive and nine negative items). In the study sample, with a total amount of 22,509 valid filled in Internet questionnaires, there were 16 occasions (0.07%) that adolescents had all 1s or 9s on the 12 emotion items. Thus, we are confident that, overall, adolescents provided relevant data.

For the analysis of the developmental trajectories, we calculated average yearly scores for each of the emotions. First, we took the sum of the three items per emotion per day and then calculated a weekly mean score. This mean score was only calculated if there were at least three valid assessment days per week. For the analyses, we used the mean of the 3 Internet assessment weeks per year (i.e., five yearly emotion scores per adolescent emotion; possible range of scores: 3–27).

Statistical Analyses

Descriptive statistics were analyzed using SPSS Version 21. Measurement invariance and developmental trajectories analyses were conducted in *Mplus* Version 6.0 (Muthén & Muthén, 1998–2010). Model fit was determined by the comparative fit index (CFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). Hu and Bentler (1999) recommend a value close to 0.95 for CFI, 0.08 for SRMR, and 0.06 for RMSEA. However, it has been suggested that these criteria might be too strict, especially when models are complex (Cheung & Rensvold, 2002). We adopted a study criterion of 0.90 or higher for CFI and 0.10 or lower for SRMR and RMSEA. Maximum likelihood estimation with robust standard errors was used to control for non-normal distributions and missing data.

Measurement Invariance Analyses

Model building. We first tested (a) measurement invariance across sex, (b) then for days within weeks, and, finally, (c) for days across the 5-year period. To answer our research questions and to test for measurement invariance, we fitted a series of hierarchically nested models (for more information on invariance testing, see Chen et al., 2005; Vandenberg & Lance, 2000).

First, we tested for configural invariance (equal factor structure across sex/time), which simultaneously served as our base models. The first set of models were multiple group (boys vs. girls), second-order factor models for each assessment day (75 models in total), to test *measurement invariance across sex* (see Figure 1 for an example). In these models, the three emotion items loaded on the associated emotion factor (e.g., "glad," "happy," and "cheerful" on Happiness; first-order factors). The four emotion factors, in turn, loaded on a general mood factor per day (i.e., Happiness, Anger, Anxiety, and Sadness were loaded on Mood; second-order factors). The factor structure was the same for boys and girls.

The second set of models consisted of longitudinal, second-order factor models for each assessment week (15 models in total), to test *measurement invariance for days within weeks* (see Figure 2 for an example). In these models, the three emotion items loaded on the associated emotion factor per day (e.g., "glad" Monday, "happy" Monday, and "cheerful" Monday on Happiness Monday; first-order factors). The daily emotion factors, in turn, loaded on a general mood factor per day (e.g., Happiness Monday, Anger Monday, Anxiety Monday, and Sadness Monday loaded on Mood Monday; second-order factors). This factor structure was the same for each day of the week.

The third set of models consisted of longitudinal, second-order factor models (15 models in total), to test *measurement invariance for days across years*. Because it was not feasible to test for invariance across all 75 days at once, we randomly selected 1 day per year (sampling drawing without replacement) and repeated the analyses with 15 different patterns (see Figure 3 for an example). In each model, 1 weekday from each of the 5 years was included (e.g., Monday Week 1 [Year 1], Tuesday Week 5 [Year 2], Wednesday Week 9 [Year 3], Thursday Week 10 [Year 4], Friday Week 15 [Year 5]). Each of the 75 days was included in one of the 15 models. The combination of the different days can be found in Online Resource 3 in the Supplemental Materials. The factor structure was the same as in the models for days within weeks (i.e., emotion items loading on emotion factors, which in turn loaded on general mood factors).

We made the following model specifications: (a) For model identification purposes, the first item's factor loading was set to 1 to set the scale of each factor and the first item's intercept was set to 0 to set the mean of each factor (Vandenberg & Lance, 2000). (b) We allowed for correlations between higher order factors (see Figures 2 and 3 for details) and for autocorrelations between residual variances of like variables over time, to account for shared method variance (e.g., item "glad" of the different days were allowed to be correlated; Vandenberg & Lance, 2000). (c) In each model, happiness was reverse coded to achieve the same scaling as in sadness, anger, and anxiety (i.e., high values indicating negative emotions).

Decision rules. Configural invariance was regarded as being established if the specified model, with the same factor structure across sex/time, fitted well (i.e., CFI \geq .90 and RMSEA/SRMR \leq 0.10). If necessary, we included correlations between residual variances as proposed by modification indices (Byrne, Shavelson, & Muthén, 1989). After having established configural invariance, we continued testing for the other measurement invariance types, namely metric invariance (equal factor loadings across sex/time), scalar invariance (equal intercepts across sex/time), and lastly strict invariance (equal residual variances across sex/time). A certain type of measurement invariance was

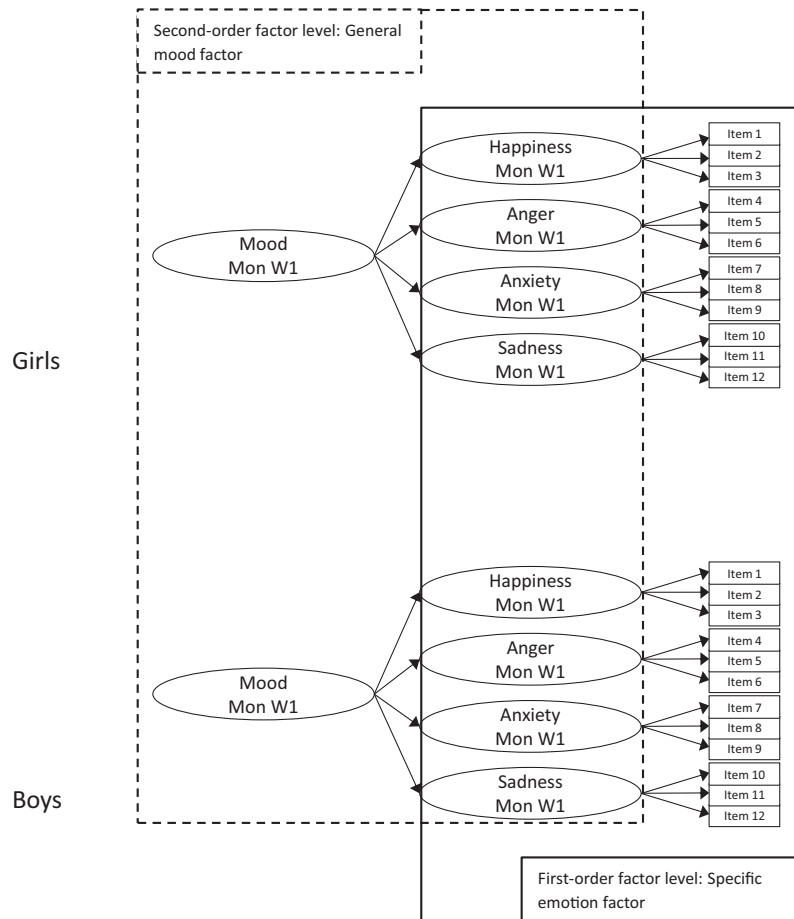


Figure 1. Multiple group measurement model for testing measurement invariance across sex (depicted for Monday in Week 1). There were 75 models tested in total. Specific emotion items load on specific emotion factors (first-order factors). Specific emotion factors load on general mood factors (second-order factor). Invariance was tested for the first-order and the second-order factors. Mon = Monday; W = Week.

established if constraining certain parameters (e.g., factor loadings in the metric invariance model) did not lead to a significant deterioration in model fit. A significant deterioration in model was indicated by a decrease greater than 0.01 in CFI (Cheung & Rensvold, 2002), which has been reported to be a reliable criterion for measurement invariance model comparisons (Chen, 2007; Cheung & Rensvold, 2002). Other fit indices that are available in *Mplus* are reported to work less well (i.e., SRMR, RMSEA, Tucker–Lewis index; Chen, 2007; Cheung & Rensvold, 2002). The chi-square difference test was not used, because it is affected by sample size, number of constraints, and nonnormality, and might be overly sensitive to trivial deviations from invariance (Chen, 2007; Cheung & Rensvold, 2002; Kelloway, 1995; Little et al., 2007). If a certain type of measurement invariance was not established, we stopped testing for subsequent types of measurement invariance.

Given that the measurement models required second-order factor models to constitute a general mood factor (see Figures 1–3), we extended the invariance testing procedure according to the recommendations of Chen and colleagues (2005). Specifically, metric, scalar, and strict invariance need to be tested on the

first-order factor level, as well as on the second-order factor level, which results in testing seven instead of four models (see Table 1 for a summary of the steps).

Developmental trajectories. To answer our research question concerning the developmental trajectories of adolescent daily emotional experiences from ages 13 to 18, we fitted latent growth models, separately for each of the four emotions (happiness, anger, sadness, and anxiety), using the yearly mean emotion scores. In latent growth models, trajectories are described by latent growth factors: the intercept (i.e., mean starting level), and linear and possibly quadratic and cubic slope factors (i.e., [non]linear change across time). For each of the four emotions, we first determined which growth curve fit our data best. Additional slope factors were only included if they were associated with a significant model fit improvement (i.e., $\Delta\text{CFI} > 0.01$). After having determined the best fitting model, we tested for sex differences in the developmental trajectories by regressing the growth factors (i.e., intercept and slopes) on adolescent sex (0 = male; 1 = female). In these models, we included SES (0 = low SES; 1 = medium/high SES) and risk status (0 = not at risk, 1 = at risk, according to TRF screening,

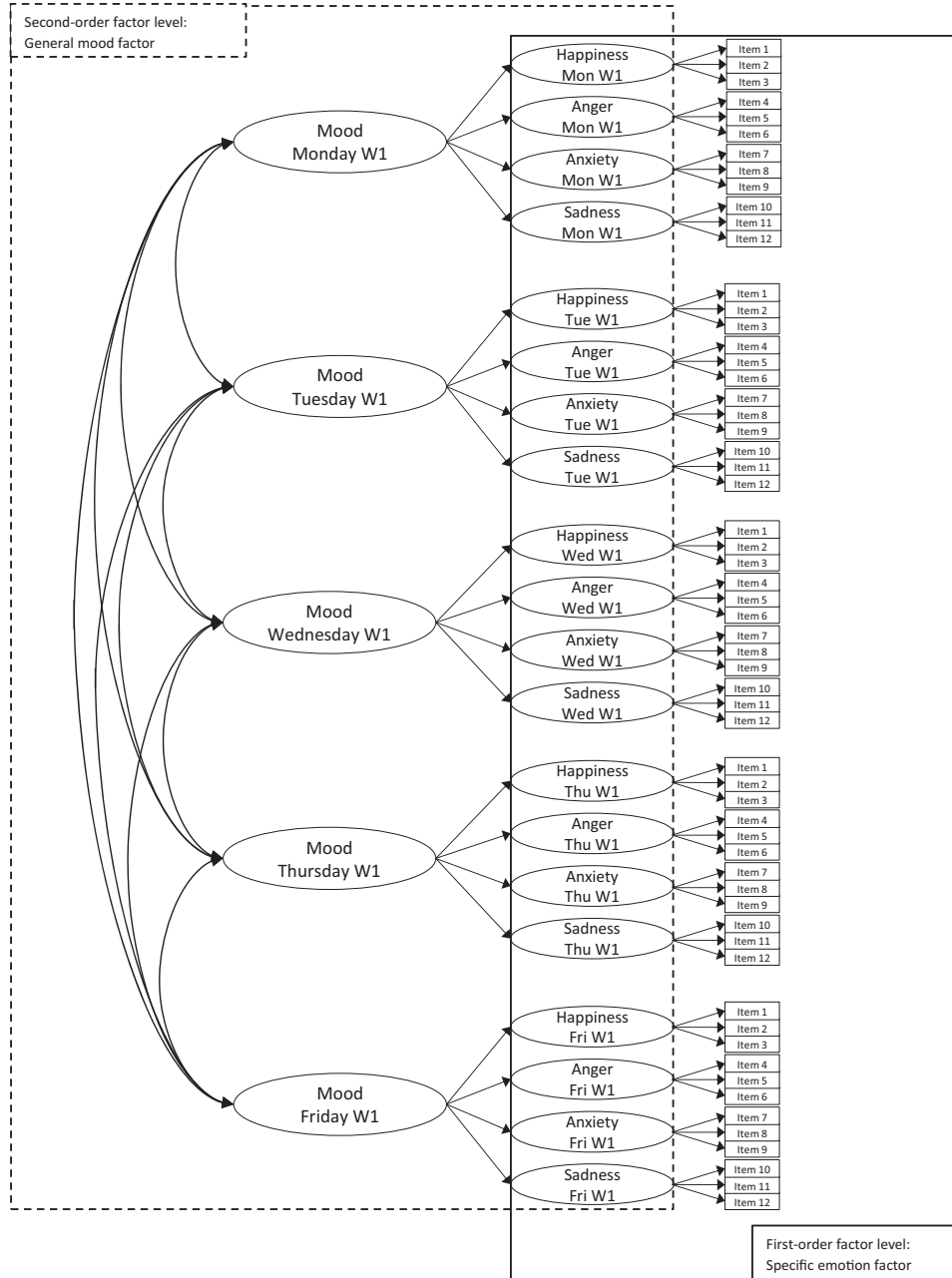


Figure 2. Measurement model for testing measurement invariance for days within a week (depicted for week 1). Note: 15 models tested in total. Autocorrelations between like variables over time were included. Specific emotion items load on specific emotion factors (first-order factors). Specific emotion factors load on general mood factors (second-order factor). Invariance was tested for the first-order and the second-order factors. Mon = Monday, Tue = Tuesday, Wed = Wednesday, Thu = Thursday, Fri = Friday, W = Week.

explained before) as time-invariant covariates by regressing the growth factors on them.

Results

Descriptive Statistics

Means and standard deviations for the items, averaged per year, as well as factor loadings for the items and the specific emotions

can be found in Table 2. Rank-order stability of the yearly emotion measures, which refer to the extent to which individuals keep their relative place across time compared to other individuals (i.e., correlations between consecutive measures of the emotion), were high, ranging from .66 to .86, all $ps < 0.001$. Moreover, results indicated that the interindividual stability was significantly higher in late than in early adolescence for all four emotions (i.e., $r_{Year4 - Year5} > r_{Year1 - Year2}$), all $ps < 0.04$.

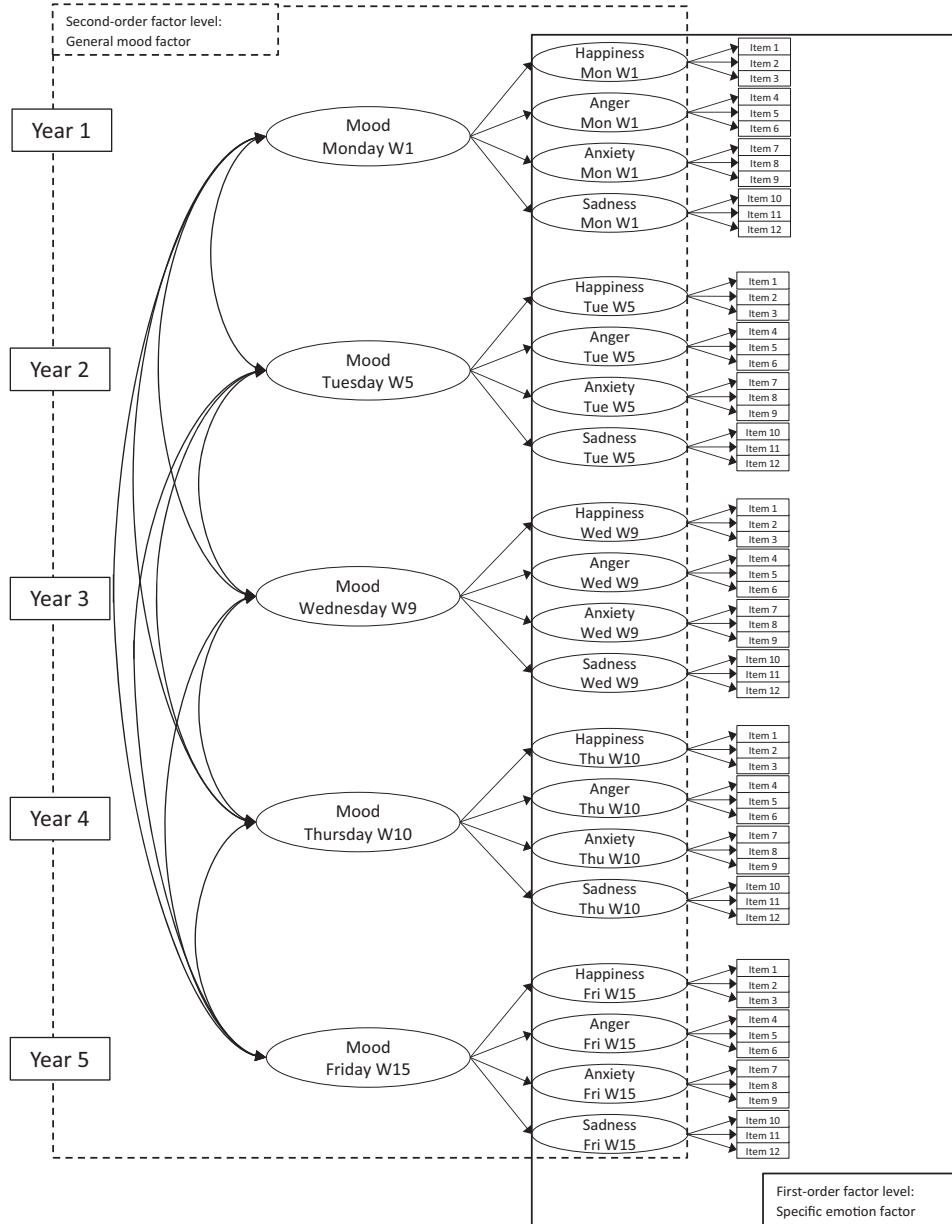


Figure 3. Measurement model for testing measurement invariance for days across years (depicted for Monday week 1, Tuesday week 5, Wednesday week 9, Thursday week 10, Friday week 15). *Note:* 15 models tested in total. For each model, we randomly picked one day per year. Each model contained a Monday, Tuesday, Wednesday, Thursday, and Friday. Autocorrelations between like variables over time were included. Specific emotion items load on specific emotion factors (first-order factors). Specific emotion factors load on general mood factors (second-order factor). Invariance was tested for the first-order and the second-order factors. Mon = Monday, Tue = Tuesday, Wed = Wednesday, Thu = Thursday, Fri = Friday, W = Week.

Measurement Invariance Analyses

Given that we analyzed a large amount of models for our measurement invariance analyses (75 models across sex, 15 models for days within weeks, 15 models for days across years), only a summary of fit statistics and model comparisons is reported (see Tables 3 and 4). The results of each model from these three model

sets can be found in the supplementary material (see Online Resources 1–3 in the Supplemental Materials). We also reanalyzed part of our models with two different inclusion criteria, to see whether different missing data patterns affected the analyses. Conclusions were, by and large, the same (see Online Resource 4 in the Supplemental Materials).

Table 1
Levels of Measurement Invariance for Second-Order Factor Models

Level of invariance	Description	
	First-order factor: Specific emotions	Second-order factor: General mood
Configural	Model 1: Same factor structure across sex/time	
Metric	Model 2: Invariance of first-order factor loadings	Model 3: Invariance of second-order factor loadings
Scalar	Model 4: Invariance of intercepts of items	Model 5: Invariance of intercepts of first-order factors
Strict	Model 7: Invariance of residual variances of items	Model 6: Invariance of disturbances of first-order factors

Note. For the first-order factors, specific emotion items loaded on specific emotion factors. For the second-order factors, specific emotions loaded on general mood factors. Usually invariance is tested first on the first-order and then on the second-order factor level. For strict invariance, tests are conducted on the second-order factor level first and then on the first-order factor level, because the former has greater theoretical interest (Chen et al., 2005).

Measurement invariance across sex. When testing measurement invariance across sex, the configural models (see Figure 1 for the factor structure), provided an acceptable fit to the data, implying that configural invariance was established across sex. Although the RMSEA was $> .10$ for 17 of the 75 models (but in no cases $> .11$), the SRMR and CFI were acceptable in these models. Metric and scalar invariance for the specific emotion and general mood factors could be established for a large majority of the models (68 of 75 models). Additional strict invariance for both factors could only be established for 21 models. Thus, we concluded that the evidence for strict invariance across sex was not convincing.

Measurement invariance for days within weeks. When testing measurement invariance for days within weeks, the configural models provided an acceptable fit to the data (see Figure 2 for factor structure), implying that configural invariance was established across days within weeks (i.e., Monday to Friday). Similarly, our analyses provided support for metric and scalar invariance for the specific emotion and general mood factors for all 15

models. Additional strict invariance for both factors was established for 13 models.

Measurement invariance for days across the 5-year period. When testing measurement invariance for days across the 5-year period, the configural models provided an acceptable fit to the data (see Figure 3 for the factor structure), implying that configural invariance was established for days across years. There was evidence for metric and scalar invariance for the specific emotion and the general mood factors for all 15 models. Additional strict invariance for both levels was only established for 9 models, which we deemed as not convincing.

Developmental Trajectories

Because there was evidence for configural, metric, and scalar measurement invariance across sex, short-term and long-term periods, we continued to test for mean level differences in the daily emotions. The model selection process revealed that adding a linear and a quadratic factor significantly improved the model fit

Table 2
Means and Standard Deviations of Items, as Well as Factor Loadings of Items and Specific Emotions

	Year 1 <i>M (SD)</i>	Year 2 <i>M (SD)</i>	Year 3 <i>M (SD)</i>	Year 4 <i>M (SD)</i>	Year 5 <i>M (SD)</i>	Factor loadings ^a
Happiness						.65
"Glad"	7.33 (1.81)	7.12 (1.69)	6.99 (1.67)	6.92 (1.61)	6.91 (1.60)	.95
"Happy"	7.33 (1.79)	7.09 (1.70)	6.97 (1.70)	6.86 (1.66)	6.84 (1.66)	.97
"Cheerful"	6.92 (1.99)	6.70 (1.93)	6.66 (1.90)	6.55 (1.88)	6.52 (1.89)	.81
Anger						.82
"Angry"	2.33 (1.84)	2.43 (1.82)	2.42 (1.75)	2.28 (1.68)	2.22 (1.64)	.91
"Cross"	2.12 (1.76)	2.22 (1.76)	2.18 (1.67)	2.09 (1.61)	2.02 (1.56)	.97
"Short-tempered"	1.88 (1.58)	1.99 (1.61)	2.00 (1.57)	1.93 (1.53)	1.87 (1.46)	.87
Anxiety						.87
"Afraid"	1.83 (1.54)	2.10 (1.68)	2.11 (1.65)	2.12 (1.65)	2.06 (1.63)	.94
"Anxious"	1.73 (1.45)	1.97 (1.60)	1.97 (1.56)	1.98 (1.56)	1.95 (1.55)	.94
"Worried"	2.42 (2.02)	2.58 (2.02)	2.59 (1.99)	2.63 (2.02)	2.64 (2.04)	.72
Sadness						.92
"Sad"	2.06 (1.80)	2.19 (1.80)	2.20 (1.75)	2.16 (1.73)	2.13 (1.73)	.91
"Down"	1.89 (1.67)	2.05 (1.70)	2.04 (1.64)	2.02 (1.63)	2.00 (1.63)	.97
"Dreary"	1.79 (1.57)	1.95 (1.63)	1.96 (1.57)	1.94 (1.57)	1.92 (1.57)	.94

Note. ^a Factor loadings for items refer to the standardized loadings of the items on the specific emotion scores (e.g., glad, happy, cheerful on happiness). Factor loadings for the specific emotion scores refer to standardized loadings of the specific emotion scores on the general negative mood scores (i.e., happiness [recoded], anger, anxiety, and sadness on negative mood). Factor loadings were obtained from the confirmatory factor analysis configural models across weeks and averaged across the study period. Means and standard deviations of the items are the averages of the means and the standard deviations of the 15 assessment days per year.

Table 3
Summary of Fit Statistics for Configural Models, Separately for Measurement Invariance Across Sex, Days Within Weeks, and Days Across Years

	Boys vs. girls ^a		Days within weeks ^b		Days across years ^c	
	Average χ^2	Average <i>df</i>	Average χ^2	Average <i>df</i>	Average χ^2	Average <i>df</i>
	219.49	100	2899.84	1520	2821.07	1522
	Range	Mean	Range	Mean	Range	Mean
CFI	.91–.97	.93	.90–.94	.91	.91–.94	.92
SRMR	.03–.09	.06	.05–.08	.06	.06–.07	.06
RMSEA	.06–.11	.09	.04–.06	.05	.04–.05	.05

Note. CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation. See Online Resources 1–3 in the Supplemental Materials for the fit indices of each model.

^a Total of 75 models. ^b Total of 15 models. ^c Total of 15 models. χ^2 and *df* are averaged across the models.

for each of the four emotions ($\Delta\text{CFI} \geq 0.05$ for linear factors, $\Delta\text{CFI} \geq 0.03$ for quadratic factors). Although additional cubic factors were associated with further model fit improvements ($\Delta\text{CFI} \geq 0.01$), they resulted in model estimation problems for all four emotion models. Therefore, we decided for the more parsimonious quadratic models. Table 5 shows the model fit indices and the means and variances of the growth parameters of the unconditional models. All models showed an acceptable model fit. Figure 4 depicts the average developmental courses of the different emotions. Happiness sharply decreased in early adolescence. This decrease leveled off by the end of adolescence. Anger slightly increased in the beginning of adolescence, but then declined again toward the end of adolescence. Sadness showed a constant, linear increase across adolescence. Anxiety increased across adolescence, with a leveling off of this increase toward late adolescence. All growth factors had significant variances, indicating that there were significant interindividual differences in the initial level and rate of change across adolescence.

There were two significant sex differences in the trajectories. First, adolescent sex had a significant influence on the linear slope of happiness ($B = -0.55, SE = 0.26, p = .04$), indicating that girls declined more steeply than boys in happiness. Second,

adolescent sex significantly predicted the intercept of sadness ($B = 0.80, SE = 0.35, p = .02$), indicating that girls had a higher initial level of sadness than boys, a difference that was stable across adolescence.

Discussion

Using daily mood data from 394 adolescents that were followed from age 13 to age 18, we tested for measurement invariance of daily mood assessments across sex, days within school weeks (i.e., short-term periods), and days across years (i.e., long-term periods). We found strong evidence for configural, metric, and scalar invariance across sex, short-term and long-term periods. Strict invariance could partially be established for days within weeks, but less convincingly across sex and days across years. Further developmental analyses indicated that happiness decreased toward middle adolescence, whereas anger, sadness, and anxiety increased. Anger returned to baseline toward late adolescence. The decrease of happiness and the increase of anxiety ceased without reversing, whereas sadness continued to increase.

Table 4
Models for Which Certain Types of Measurement Invariance Were Supported, Separately for Models Across Sex, Days Within Weeks, Days Across Years

Type of Invariance	(A) Boys vs. girls		(B) Days within weeks		(C) Days across years	
	First-order factor	Second-order factor	First-order factor	Second-order factor	First-order factor	Second-order factor
Step 1: Configural (same pattern of factor loadings across sex/time)	75/75 models		15/15 models		15/15 models	
Step 2: Metric (+ equal factor loadings across sex/time)	73/75 models	72/73 models	15/15 models	15/15 models	15/15 models	15/15 models
Step 3: Scalar (+ equal intercepts across sex/time)	68/72 models	68/68 models	15/15 models	15/15 models	15/15 models	15/15 models
Step 4: Strict (+ equal residual variances across sex/time)	21/56 models	56/68 models	13/15 models	15/15 models	9/15 models	15/15 models

Note. Table displays the amount of models for which a certain kind of measurement invariance was supported in relation to the amount of models for which invariance was tested (models supported/models tested). A certain kind of measurement invariance was supported, when constraining parameters (e.g., intercepts for scalar invariance) did not lead to a worse model fit ($\Delta\text{CFI} \leq -.01$). In cases that a certain type of measurement invariance was not established for a model (e.g., Monday in Week 1), we did not test for subsequent stages of measurement invariance for that particular model (i.e., that is why the total number of tested model may diminish across the different steps of measurement invariance). First-order factors refer to specific emotions (items loading on emotion factors), second-order factors refer to general mood factors (emotions loading on general mood factors).

Table 5

Results of Unconditional Latent Growth Curve Analyses of Happiness, Anger, Sadness, and Anxiety Level

	Model fit indices					Intercept		Linear slope		Quadratic slope	
	χ^2	df	CFI	SRMR	RMSEA	Mean	Variance	Mean	Variance	Mean	Variance
Happiness ^a	26.71	7	.98	.10	.09	21.53***	10.02***	-.65***	2.78**	.06*	.13***
Anger	13.39	6	.99	.02	.06	6.39***	8.67***	.30*	3.80***	-.07*	.21***
Sadness	30.45	6	.95	.03	.10	5.84***	8.21***	.33*	2.86**	-.05	.20***
Anxiety	25.91	6	.97	.03	.09	6.02***	10.16***	.55***	4.81***	-.08*	.27***

Note. Results show unstandardized coefficients. CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation.

^a The residual variance of happiness in Year 5 was negative, but not significant ($p = .57$). It was fixed to zero to avoid problems with the latent variable covariance (PSI) matrix and allow the model to run. This modification accounts for the differences in the degrees of freedom compared to the other models. * $p < .05$. ** $p < .01$. *** $p < .001$.

Measurement Invariance Across Sex, Short-Term, and Long-Term Periods

The fact that configural and metric invariance were consistently established across sex and time indicates that the structure and unit of measurement (i.e., factor loadings) is equal between boys and

girls and different time frames. This implies that the items of the different emotions have the same importance for the emotion scores for boys and girls, on different weekdays, and different ages (e.g., item “worried” for the emotion anxiety). Moreover, it also suggests that the different emotions have the same importance for

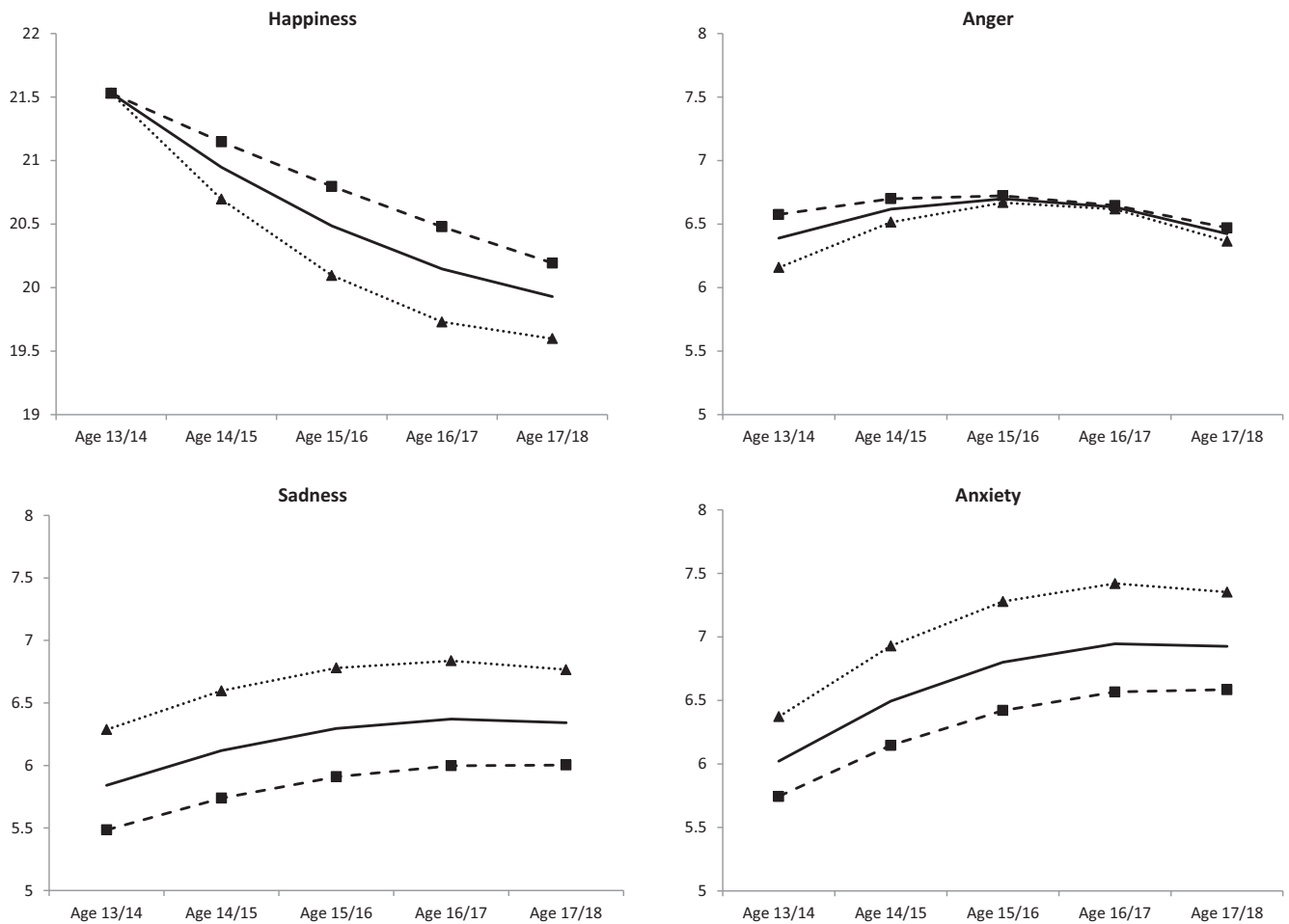


Figure 4. Developmental trajectories of daily emotions across adolescence (happiness, anger, sadness, and anxiety), for overall sample (—) and separately for males (---■) and females (---▲). Note: Risk status and SES were included as covariates. Possible range of scores: 3–27.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the general negative mood factors across sex and time. For instance, sadness contributes in an equal way to general negative mood for boys and girls, and for different weekdays and years. Due to the establishment of metric invariance, it is justified to compare associations among daily emotion ratings and relations of these with other external variables across sex, and shorter and longer time-frames.

The finding that scalar invariance was also established for all three model sets indicates that the starting point in rating emotions is equal between boys and girls and does not change across shorter and longer time frames. Scalar invariance makes it possible to compare means between boys and girls and across time. This finding implies that the developmental changes in daily emotions can be fully attributed to developmental differences and not to changes in response tendencies across adolescence. Moreover, because the emotion measurement constructs stay equal across the week, results also suggest that day-to-day mood scores can be aggregated to a more robust measure for between-person analyses, such as weekly mood level or variability scores. Recent studies have used day-to-day level and variability scores as indicators of emotional dysregulation and have shown that high negative mood levels and high mood variability are risk factors for various forms of psychopathology (for a recent meta-analysis, see Houben, Van Den Noortgate, & Kuppens, 2015; for a methodological paper, see Jahng, Wood, & Trull, 2008).

The fact that strict invariance was not convincingly established is not surprising, because it is often unrealistic that error influences do not change across time or are equal between boys and girls (Chen et al., 2005; Gregorich, 2006; Little et al., 2007). For instance, concerning sex, it simply indicates that random experiences during the day of assessment are likely to have sex-specific influences on mood ratings. However, a failure to establish strict invariance has little additional practical value over scalar invariance and is not necessary to test for differences in associations or means over time or across sex (Chen et al., 2005).

Developmental Changes in Emotional Experiences Across Adolescence

Because we found evidence for configural, metric, and scalar invariance across sex, short-term and long-term periods, we further examined developmental trajectories of average emotional experiences across adolescence. Our results point to important developmental changes in emotions during that time. The results are, overall, in line with previous studies (Larson et al., 2002; Moneta et al., 2001), which found a similar deterioration of overall mood across adolescence. One contribution of our study is that we examined trajectories of different emotions. Our results suggest that the different emotions do not entirely develop in the same way. Based on previous longitudinal research using more conventional self-report questionnaires (e.g., Karriker-Jaffe et al., 2008; Nelemans et al., 2014; Van Oort et al., 2009), we predicted initial increases in anger, followed by decreases, increases in sadness, and decreases in anxiety, followed by increases. We did not form predictions about happiness due to limited previous research. Our results were largely in line with these predictions. We found that anger peaked during middle adolescence and that sad emotions continuously increased across adolescence. Anxiety did not show the initial decrease followed by an increase, as hypothesized, but

increased already in early adolescence with a leveling off of this increase toward late adolescence (see also Nelemans et al., 2014). Happiness decreased across adolescence with a slight leveling off toward late adolescence, which is in line with the increase in anxiety and sadness. To summarize, our daily diary results of different emotions largely converge with results from other studies which used more state-like assessments of emotions or psychopathological symptoms, thus contributing to the robustness of such findings.

It is interesting that anger returned to baseline at the end of adolescence, which is in contrast to the development of happiness, sadness, and anxiety, which stayed low and high, respectively. This is in line with a recent study spanning from early to late adolescence that showed that direct aggressive behavior peaked in middle adolescence, whereas generalized anxiety continuously increased (Meeus et al., 2016). Anger may decrease toward late adolescence, because high anger levels can jeopardize everyday functioning in adulthood (e.g., in romantic relationships or in job environments). With regard to the other emotions, it seems that the more positive mood from the beginning of adolescence has disappeared by the end of adolescence. One reason for that might be emerging adulthood, in which individuals have to become independent and take responsibility for themselves (Arnett, 2000). This might be accompanied by such negatively tuned anxious and sad emotions. However, during emerging adulthood, moods seem to become more positive again, as indicated by a study that showed increases in self-esteem and decreases in anger and depression levels from ages 18 to 25 (Galambos, Barker, & Krahn, 2006).

There was a lot of interindividual variability in the developmental trajectories of the different emotions, as indicated by the significant variances of the intercept and slope factors. This suggests that adolescents differ in their overall emotion level (e.g., some adolescents are in general sadder than others) as well as in their developmental change across time (e.g., some adolescents may show steeper increases in sadness or level off to a lesser extent than others). The rank-order stability was quite high and significantly increased toward late adolescence. This indicates that a large majority of adolescents keep their relative place across time, especially when they get older (i.e., an adolescent with high levels of sadness at one time is also likely to report high levels at another time).

Our results further pointed toward some sex differences. Previous studies that merged positive and negative moods found little evidence for sex-specific developments of average moods (Larson et al., 2002; Moneta et al., 2001). Our study suggests that this is largely true. The only differences we found were that girls reported more sadness in general and experienced steeper declines in happiness compared to boys. This difference might be explained by studies that show that girls use more rumination strategies than boys, which in turn has been linked to higher levels of daily sadness (Silk et al., 2003). It is intriguing that girls showed steeper declines in happiness than boys, whereas we did not find such patterns for the other emotions, especially not for sadness. This might suggest that the increase of depressive symptoms that has been frequently reported for girls (for a review, see Hankin & Abramson, 2001) might be driven by a more dramatic decline in happiness rather than a larger increase in sadness, of which girls already showed higher levels in early adolescence. Moreover, the results suggest that greater decreases in positive mood do not

necessarily coincide with greater increases in negative mood and provide further evidence that positive and negative are separate constructs (Lonigan et al., 1999).

Limitations, Future Directions, and Conclusions

A number of limitations need to be mentioned when interpreting the results. First, our sample included mostly high to middle-class, Dutch adolescents. More than one third of our participants had borderline or higher scores on externalizing behavior at age 12. This might have affected the generalizability of our findings. It is unclear whether our results would generalize to samples with more diverse ethnical, socioeconomic, or psychopathological characteristics.

Second, mood ratings were only available from Monday until Friday in the present study. Due to the omission of Saturdays and Sundays, it is not clear whether these results would generalize to weekends. Including Saturdays and Sundays is an important direction for future research, particularly because studies have shown that adolescents experience significantly more positive moods on weekends than on weekdays (e.g., Larson & Richards, 1998). Similarly, although the present study was the first to study developmental trajectories of four primary emotions using daily diary data, it omitted secondary (e.g., shame, pride) and background (e.g., tiredness) emotions. A direction for future research is thus to also include these emotions and to study their developmental trajectories as well as their (longitudinal) relation with primary emotions.

Third, the extensive daily diary design in our study over the course of 5 years, with three assessment weeks per year, makes missing data inevitable and thus only a subset of adolescents was included in the main analyses. To address this, we replicated part of our results across different missing data patterns and showed that emotion ratings did not differ between adolescents that dropped out compared to adolescents that did not drop out. Moreover, in the overall sample, 80% of adolescents were still participating the last year of the study. This is comparable to similar studies in adolescents (Connell, Dishion, & Klostermann, 2012; Ormel et al., 2012).

Lastly, mood was measured only once per day in the present study. It has been noted that recall biases may happen even within a short time frame, and that study participants may find it difficult to summarize their mood for one day (Shiffman et al., 2008). However, a recent meta-analysis found that relations between emotional variability scores and well-being were consistent across different time-scales of the emotion data collection (intervals ranging from seconds, minutes, hours, and up to days; Houben et al., 2015). Moreover, another study showed that between-day and within-day emotion variability scores are highly correlated ($r = .75$; Jahng et al., 2008). Nevertheless, the present study cannot draw conclusions on measurement invariance for smaller time intervals.

Despite these limitations, results of the present study suggest that differences in emotion ratings between boys and girls, between weekdays, and between years can be adequately interpreted. This is particularly interesting because it indicates that the structure of emotions, as collected with daily diaries in the present study, does not differ between boys and girls and also does not change across shorter and longer time-frames. Moreover, our re-

sults suggest that adolescence is a time in which adolescents become less happy and that it is important to consider different emotions to uncover developmental or sex differences. The results could be illuminating for our understanding of psychopathological phenomena during adolescence, in which emotions play a central role (Cole & Hall, 2008). Due to the high interindividual variability in the trajectories, future studies should examine whether subtypes of emotional trajectories exist, for instance with growth mixture modeling techniques (see Nelemans et al., 2014 for an example on anxiety symptom trajectories). It is important to note that our analyses do not imply that all daily emotion data collected in adolescents are measurement invariant across sex and time. Like other reliability and validity studies, these analyses only apply to this specific sample, utilizing this specific timing, and content of measurement. Although it can be a daunting task to establish measurement invariance of a construct with such complex longitudinal daily diary data, we want to encourage researchers to follow our example to gain confidence in their results on interindividual differences and longitudinal change. As Widaman et al. (2010) put it, “[Measurement invariance is] one of the more vexing problems in assessing development—and one that deserves greater attention” (p. 10).

References

- Achenbach, T. M. (1991). *Manual for the Youth Self Report and 1991 profiles*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist, 54*, 317–326. <http://dx.doi.org/10.1037/0003-066X.54.5.317>
- Arnett, J. J. (2000). Emerging adulthood. A theory of development from the late teens through the twenties. *American Psychologist, 55*, 469–480. <http://dx.doi.org/10.1037/0003-066X.55.5.469>
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466. <http://dx.doi.org/10.1037/0033-2909.105.3.456>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. <http://dx.doi.org/10.1080/10705510701301834>
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher’s corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471–492. http://dx.doi.org/10.1207/s15328007sem1203_7
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5
- Cole, D. A., Tram, J. M., Martin, J. M., Hoffman, K. B., Ruiz, M. D., Jacquez, F. M., & Maschman, T. L. (2002). Individual differences in the emergence of depressive symptoms in children and adolescents: A longitudinal investigation of parent and child reports. *Journal of Abnormal Psychology, 111*, 156–165. <http://dx.doi.org/10.1037/0021-843X.111.1.156>
- Cole, P. M., & Hall, S. E. (2008). Emotion dysregulation as a risk factor for psychopathology. In T. P. Beauchaine & S. P. Hinshaw (Eds.), *Child and adolescent psychopathology* (pp. 265–298). Hoboken, NJ: Wiley.
- Connell, A. M., Dishion, T. J., & Klostermann, S. (2012). Family check up effects on adolescent arrest trajectories: Variation by developmental subtype. *Journal of Research on Adolescence, 22*, 367–380. <http://dx.doi.org/10.1111/j.1532-7795.2011.00765.x>
- Cox, D. L., Stabb, S. D., & Hulgus, J. F. (2000). Anger and depression in girls and boys: A study of gender differences. *Psychology of Women*

- Quarterly*, 24, 110–112. <http://dx.doi.org/10.1111/j.1471-6402.2000.tb01027.x>
- Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, self-esteem, and anger in emerging adulthood: Seven-year trajectories. *Developmental Psychology*, 42, 350–365. <http://dx.doi.org/10.1037/0012-1649.42.2.350>
- Ge, X., Lorenz, F. O., Conger, R. D., Elder, G. H., & Simons, R. L. (1994). Trajectories of stressful life events and depressive symptoms during adolescence. *Developmental Psychology*, 30, 467–483. <http://dx.doi.org/10.1037/0012-1649.30.4.467>
- Gjerde, P. F. (1995). Alternative pathways to chronic depressive symptoms in young adults: Gender differences in developmental trajectories. *Child Development*, 66, 1277–1300. <http://dx.doi.org/10.2307/1131647>
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44, S78–S94. <http://dx.doi.org/10.1097/01.mlr.0000245454.12228.8f>
- Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., . . . Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21. *Neuropsychology*, 26, 251–265. <http://dx.doi.org/10.1037/a0026712>
- Hankin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. *Psychological Bulletin*, 127, 773–796. <http://dx.doi.org/10.1037/0033-2909.127.6.773>
- Hoeksma, J. B., Sep, S. M., Vester, F. C., Groot, P. F., Sijmons, R., & De Vries, J. (2000). The electronic mood device: Design, construction, and application. *Behavior Research Methods*, 32, 322–326. <http://dx.doi.org/10.3758/BF03207801>
- Holmbeck, G. N., Friedman, D., Abad, M., & Jandasek, B. (2006). Development and psychopathology in adolescence. In D. A. W. E. J. Mash (Ed.), *Behavioral and Emotional Disorders in Adolescents: Nature, Assessment, and Treatment* (pp. 21–55). New York, NY: Guilford Press Publications.
- 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. <http://dx.doi.org/10.1037/a0038822>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13, 354–375. <http://dx.doi.org/10.1037/a0014173>
- Karriker-Jaffe, K. J., Foshee, V. A., Ennett, S. T., & Suchindran, C. (2008). The development of aggression during adolescence: Sex differences in trajectories of physical and social aggression among youth in rural areas. *Journal of Abnormal Child Psychology*, 36, 1227–1236. <http://dx.doi.org/10.1007/s10802-008-9245-5>
- Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior*, 16, 215–224. <http://dx.doi.org/10.1002/job.4030160304>
- Kemper, T. D. (1987). How many emotions are there? Wedding the social and the autonomic components. *American Journal of Sociology*, 93, 263–289. <http://dx.doi.org/10.1086/228745>
- Khoo, S.-T., West, S. G., Wu, W., & Kwok, O.-M. (2006). Longitudinal methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 301–317). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/11383-021>
- Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, 7, 297–300. <http://dx.doi.org/10.1177/1754073915590947>
- Larson, R. W., & Ham, M. (1993). Stress and “storm and stress” in early adolescence: The relationship of negative events with dysphoric affect. *Developmental Psychology*, 29, 130–140. <http://dx.doi.org/10.1037/0012-1649.29.1.130>
- Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, 73, 1151–1165. <http://dx.doi.org/10.1111/1467-8624.00464>
- Larson, R., & Richards, M. (1998). Waiting for the weekend: Friday and Saturday night as the emotional climax of the week. *New Directions for Child and Adolescent Development*, 82, 37–51. <http://dx.doi.org/10.1002/cd.23219988204>
- Larson, R. W., Richards, M. H., & Perry-Jenkins, M. (1994). Divergent worlds: The daily emotional experience of mothers and fathers in the domestic and public spheres. *Journal of Personality and Social Psychology*, 67, 1034–1046. <http://dx.doi.org/10.1037/0022-3514.67.6.1034>
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357–365. <http://dx.doi.org/10.1177/0165025407077757>
- Lonigan, C. J., Hooe, E. S., David, C. F., & Kistner, J. A. (1999). Positive and negative affectivity in children: Confirmatory factor analysis of a two-factor model and its relation to symptoms of anxiety and depression. *Journal of Consulting and Clinical Psychology*, 67, 374–386. <http://dx.doi.org/10.1037/0022-006X.67.3.374>
- Meeus, W., Van de Schoot, R., Hawk, S. T., Hale, W. W., III, & Branje, S. (2016). Direct aggression and generalized anxiety in adolescence: Heterogeneity in development and intra-individual change. *Journal of Youth and Adolescence*, 45, 361–375. <http://dx.doi.org/10.1007/s10964-015-0388-8>
- Moneta, G. B., Schneider, B., & Csikszentmihalyi, M. (2001). A longitudinal study of the self-concept and experiential components of self-worth and affect across adolescence. *Applied Developmental Science*, 5, 125–142. http://dx.doi.org/10.1207/S1532480XADS0503_2
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide, sixth edition*. Los Angeles, CA: Authors.
- Nelemans, S. A., Hale, W. W. I., Branje, S. J. T., Raaijmakers, Q. A. W., Frijns, T., van Lier, P. A. C., & Meeus, W. H. J. (2014). Heterogeneity in development of adolescent anxiety disorder symptoms in an 8-year longitudinal community study. *Development and Psychopathology*, 26, 181–202. <http://dx.doi.org/10.1017/S0954579413000503>
- Neumann, A., van Lier, P. A. C., Frijns, T., Meeus, W., & Koot, H. M. (2011). Emotional dynamics in the development of early adolescent psychopathology: A one-year longitudinal study. *Journal of Abnormal Child Psychology*, 39, 657–669. <http://dx.doi.org/10.1007/s10802-011-9509-3>
- Neumann, A., van Lier, P. A. C., Gratz, K. L., & Koot, H. M. (2010). Multidimensional assessment of emotion regulation difficulties in adolescents using the Difficulties in Emotion Regulation Scale. *Assessment*, 17, 138–149. <http://dx.doi.org/10.1177/1073191109349579>
- Ormel, J., Oldehinkel, A. J., Sijtsma, J., van Oort, F., Raven, D., Veenstra, R., . . . Verhulst, F. C. (2012). The Tracking Adolescents' Individual Lives Survey (TRAILS): Design, current status, and selected findings. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51, 1020–1036. <http://dx.doi.org/10.1016/j.jaac.2012.08.004>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents' emotion regulation in daily life: Links to depressive symptoms and problem

- behavior. *Child Development*, 74, 1869–1880. <http://dx.doi.org/10.1046/j.1467-8624.2003.00643.x>
- Sterba, S. K., Copeland, W., Egger, H. L., Jane Costello, E., Erkanli, A., & Angold, A. (2010). Longitudinal dimensionality of adolescent psychopathology: Testing the differentiation hypothesis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 51, 871–884. <http://dx.doi.org/10.1111/j.1469-7610.2010.02234.x>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. <http://dx.doi.org/10.1177/109442810031002>
- Van Oort, F. V. A., Greaves-Lord, K., Verhulst, F. C., Ormel, J., & Huizink, A. C. (2009). The developmental course of anxiety symptoms during adolescence: The TRAILS study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 50, 1209–1217. <http://dx.doi.org/10.1111/j.1469-7610.2009.02092.x>
- Weems, C. F. (2008). Developmental trajectories of childhood anxiety: Identifying continuity and change in anxious emotion. *Developmental Review*, 28, 488–502. <http://dx.doi.org/10.1016/j.dr.2008.01.001>
- Weinstein, S. M., Mermelstein, R. J., Hankin, B. L., Hedeker, D., & Flay, B. R. (2007). Longitudinal patterns of daily affect and global mood during adolescence. *Journal of Research on Adolescence*, 17, 587–600. <http://dx.doi.org/10.1111/j.1532-7795.2007.00536.x>
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4, 10–18. <http://dx.doi.org/10.1111/j.1750-8606.2009.00110.x>
- Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology*, 17, 251–257. <http://dx.doi.org/10.1016/j.conb.2007.03.009>

Received August 17, 2015

Revision received February 22, 2016

Accepted February 26, 2016 ■