AMERICAN PSYCHOLOGICAL ASSOCIATION

Emotion

ISSN: 1528-3542

2021, Vol. 21, No. 8, 1637–1649
https://doi.org/10.1037/emo0001031

Rosy or Blue? Change in Recall Bias of Students' Affective Experiences During Early Adolescence

Carmen L. A. Zurbriggen[1], David Jendryczko[2], and Fridtjof W. Nussbeck[2]
[1] Faculty of Humanities, Education and Social Sciences, University of Luxembourg
[2] Department of Psychology, University of Konstanz

Changes in the quality of emotional experience are among the various significant developmental challenges that characterize early adolescence. Although retrospective and momentary emotional self-reports are known to differ, adolescents' emotional experiences are mainly assessed retrospectively without knowing if the recall is biased in a positive or negative way. The present study extends research on recall bias by investigating possible changes in retrospection effects of students' affective experiences during early adolescence. To this end, we compared retrospectively assessed affect with in situ reported affect. At the age of about 12 years (T1) and 3 years later (T2), 120 students repeatedly reported their momentary positive and negative affect during one school week and once in retrospective at the end of the school week. Furthermore, we examined whether students' emotional attitudes toward school have an effect on retrospection effects of students' affect and on change in retrospection effects from T1 to T2. To test our hypotheses, we applied multilevel first-order and second-order latent difference models. Results indicate a positive recall bias (i.e., rosy view) of students' reports of their positive and negative affect in the classroom at T1 and a negative shift in recall bias by T2. Furthermore, findings supported that a rosy view is less likely to occur, if a student is less emotionally involved in school. In turn, positive emotional attitudes toward school appear to serve as a buffer for the tendency toward a negative recall bias (i.e., blue view) at the end of early adolescence.

Keywords: affect, recall bias, early adolescence, emotional attitudes, latent difference model

Supplemental materials: https://doi.org/10.1037/emo0001031.supp

Do adolescents report what they actually feel? Although this question may seem rather trivial, it is fundamental for investigating adolescents' emotional experiences. Since emotions, feelings, and affect are subjective phenomena, most researchers consider the individual having access to these internal states as "the final arbiter of his or her own feeling" (Schwarz, 2012, p. 31). This is also true for adolescents.

In educational research, adolescents' emotional experiences are mainly assessed by means of conventional questionnaires, researchers frequently rely on retrospective self-reports; thus, on a memory-based reporting format. However, reports of recalled emotions or experiences are often over- or underestimated (Conner & Barrett, 2012; Ottenstein & Lischetzke, 2020). Previous research comparing retrospectively assessed emotions with emotions captured in situ or in real-time has consistently revealed differences between retrospective and in situ reports (e.g., Christensen et al., 2003; Lay et al., 2017; Mill et al., 2015; Parkinson et al., 1995; Wirtz et al., 2003). These retrospection effects (i.e., recall biases) can partly be explained by the fleeting and fluctuating characteristics of emotions. Once the feeling dissipates, the emotional experience needs to be mentally reconstructed by the respondent based on the temporarily most accessible and relevant information (Robinson & Clore, 2002). Hence, the retrospective rating depends also on circumstances rendering pieces of information more or less salient.

With respect to adolescents, it is important to investigate influences on self-reports of emotional experiences in the classroom, since school represents a major developmental context for adolescents' psychological functioning (Eccles & Roeser, 2011; Moksnes et al., 2016). Following Robinson and Clore's (2002) accessibility model of emotional self-reports, it is assumed that adolescents access situation-specific beliefs when retrospectively rating their emotional experiences. The few studies on the bias of students' emotional self-reports in the classroom indicate that adolescents' retrospective or prospective assessment

This article was published Online First December 20, 2021.
Carmen L. A. Zurbriggen https://orcid.org/0000-0002-0906-2000
David Jendryczko https://orcid.org/0000-0002-5984-8765
Fridtjof W. Nussbeck https://orcid.org/0000-0003-4002-8916
We have no known conflict of interest to disclose. This article is based on data from a project funded by the Swiss National Research Foundation (Grant 122274) and the University of Applied Sciences of Special Needs Education Zurich, Switzerland. We commemorate Martin Venetz, who passed away in May 2018, as first author of a conference presentation that served as basis for this article.
Correspondence concerning this article should be addressed to Carmen L. A. Zurbriggen, Faculty of Humanities, Education and Social Sciences, University of Luxembourg, Campus Belval, 11 Portes des Sciences, 4366 Esch-sur-Alzette, Luxembourg. Email: carmen.zurbriggen@uni.lu

of emotions or affect during lessons is influenced by their school-related beliefs (Bieg et al., 2014; Goetz et al., 2013; Venetz & Zurbriggen, 2016). For students in Grade 6, Venetz and Zurbriggen (2016) showed that the positive bias (i.e., rosy view) of students' retrospective affect ratings in the classroom was primarily associated with students' emotional attitudes toward school. Given the decline in the positivity of emotions (Larson et al., 2002; McLaughlin et al., 2015) and in well-being during early adolescence (Casas & González-Carrasco, 2019), as well as the increase in negative attitudes toward school (Hascher & Hadjar, 2018) and a decrease in students' motivation in secondary education (Minnaert et al., 2017; Wigfield et al., 2006), the question arises whether the positive bias of retrospective affective self-reports would remain stable or deteriorate in this sensitive period of life.

The present study extends research on recall bias by investigating possible changes in retrospection effects of students' affective experiences in the classroom during early adolescence. Furthermore, the impact of students' attitudes toward school on retrospection effects is explored. Due to the frequently identified difficulties of adolescents in emotion differentiation (Nook et al., 2018), we refrain from investigating discrete emotions and rely instead on the concept of core affect, which assumes that affective states are always potentially accessible to consciousness and experienced as a simple, nonreflective feeling (Russell, 2003; Yik et al., 2011). Core affect is well suited to investigate affective experience in daily life (Kuppens et al., 2012).

Recall Bias of Emotional Self-Reports

Emotional experience is continuous and fluctuating and does not endure as such in memory (Levine et al., 2009; Robinson & Clore, 2002; Schwarz, 2012). According to this common view in research on emotional self-reports, memory for emotions or affect is subject to change, forgetting or bias over time. Since individuals cannot retrieve the emotional experience itself, they construct a representation of the feeling based on their memory about the circumstances while experiencing the feeling, and evaluate the reconstruction based on their beliefs related to the past feeling (Levine et al., 2009). The process of recalling information can be considered as a "judgment, implying a dynamic process comprised of the differential weighting of information" (Stone & Litcher-Kelly, 2006, p. 63).

The possibility that biases influence recall ratings has driven, among others, the development of momentary data capturing techniques. Techniques for studying people in their real life cover a wide range of assessment methods, such as diary methods (e.g., Bolger et al., 2003), experience sampling method (ESM; e.g., Hektner et al., 2007), ambulatory assessment (e.g., Conner & Mehl, 2015; Trull & Ebner-Priemer, 2013), or the recently increasingly used umbrella term, intensive longitudinal methods (e.g., Bolger & Laurenceau, 2013). The main characteristics of these methods are that data are collected (a) in situ, (b) in real time or close to real time, and (c) on several occasions over a period of time (e.g., one or several weeks). With regard to self-reports of emotional experiences, one of the main advantages of momentary assessment is that recall biases can be mitigated or even greatly reduced (Stone & Litcher-Kelly, 2006) due to the direct accessibility of information (Schwarz, 2012).

Accordingly, many studies have reported differences between momentary and retrospective reports of emotions, affect or experiences in general (e.g., Christensen et al., 2003; Lay et al., 2017; Parkinson et al., 1995; Wirtz et al., 2003). To consider recall bias, most studies have used time-inclusive retrospective reports (Ottenstein & Lischetzke, 2020). With this type of retrospective report, recall bias corresponds to the difference between the average of the momentary ratings given by a participant during a specific time span and the retrospective rating about this time period overall. The difference or convergence between aggregated momentary ratings and the overall retrospective ratings has also been referred to as memory-experience gap (Miron-Shatz et al., 2009) or as (mean) level convergence (Neubauer et al., 2020). In this contribution, we explicitly use the term retrospection bias since we consider in situ ratings to better reflect emotional experiences.

Notwithstanding the differences in terminology, possible explanations for (the lack of) convergence between momentary and retrospective ratings remain an important question in emotion research. Robinson and Clore (2002) argue that different types of emotional self-report are associated with different sources of information that raters rely upon, providing potentially different answers about their emotional experience. According to their accessibility model of emotional self-report, time-inclusive retrospective reports are supposed to be linked to semantic rather than to episodic emotion knowledge. With the successive decline in a person's accessibility of episodic information (i.e., information that is specific to an event in the past), the reliance on semantic knowledge (i.e., generalized beliefs) increases. Such generalized beliefs can be either situation-specific or identity-related. When a situation or a setting is connected with particular beliefs about one's emotions, the rater more likely relies on situation-specific beliefs than on (general) identity-related beliefs.

For adolescents, it can be argued that school is associated with particular situation-specific beliefs. Accordingly, and in line with Robinson and Clore's accessibility model of emotional self-reports, adolescents may rely on school-related beliefs, that is, attitudes toward school when evaluating their affective experience retrospectively. Following the general definition of attitude by Eagly and Chaiken (2007), attitude toward school can be defined as an inner tendency related to school that is expressed by evaluating school with some degree of (dis-)favor. Because attitudes toward school are an important predictor of students' well-being (Moè et al., 2009), they might also have an impact on its retrospective (or prospective) assessment in the classroom. The importance of attitudes toward school for students' emotional self-reports can be further substantiated by the Affective Events Theory (Weiss & Cropanzano, 1996). Drawing a parallel between work context (in adult life) and school context, it can be assumed that cumulative emotional experiences are closely connected with students' school related attitudes.

Changes in Emotional Experience in Early Adolescence

Age is a main factor that can influence retrospective emotion or affect ratings. Studies investigating age differences suggest a tendency for more positivity and a reduced negativity in both affective experience and emotional memory in older age than in early adulthood (Charles et al., 2016; Lay et al., 2017; Mill et al., 2015; Röcke et al., 2011). To the best of our knowledge, there is no

study on age differences in recall bias of emotional self-reports in adolescence. However, we refer to research on emotional development to derive hypotheses about the development of recall bias in adolescents.

Changes in the quality of emotional experience and subjective well-being are among the various significant developmental changes that characterize adolescence (McLaughlin et al., 2015; Wigfield et al., 2006). Numerous studies have demonstrated an increase in mental health problems such as anxiety, depression, or suicidal thoughts (e.g., Bertha & Balázs, 2013; Valois et al., 2004) as well as a decrease in life satisfaction (e.g., Goldbeck et al., 2007; Newland et al., 2019). A large international comparative study revealed that general subjective well-being continuously decreased from age 7 to 14, with an onset of the decrease at the age of 10 in most countries (Casas & González-Carrasco, 2019). This trend was not related to school transition (i.e., from primary to secondary education).

Studies using momentary data capture techniques support the notion that the early years of adolescence mark the beginning of a downward tendency in emotional development. In an early ESM study by Larson and Ham (1993), for instance, students experienced more negative affect in Grades 7 to 9 than in Grades 5 and 6. Additionally, adolescents' emotional experiences became not only less positive, but also less stable across early adolescence with a slowing of emotional changes in middle adolescence (Larson et al., 2002). Weinstein and colleagues (Weinstein et al., 2007) found, that while positive affect declined across Grades 8 to 11, negative affect remained relatively stable. Relative stability in depressed mood (i.e., general ratings of depressive symptoms) throughout adolescence was also shown in a 6-year longitudinal panel study (Holsen et al., 2000).

According to Hektner (2014), the increase in negative emotional experiences in early adolescence may be related to a shift in the perception of negative life events (see also Larson & Ham, 1993). Another possible explanation is that adolescents may have a preference for maintaining or even enhancing negative affect at the expense of positive affect—a phenomenon called contra-hedonic motivation. Riediger and colleagues (Riediger et al., 2009) found a relatively high prevalence of contra-hedonic motivation in adolescents, but also a higher prevalence of mixed affects (i.e., the simultaneous experience of positive and negative affect of high intensity) compared with older age-groups. Furthermore, the pattern of higher negative affect and lower positive affect in early adolescence could be related to insufficient or inappropriate emotion regulation. Findings from a study by Zimmermann and Iwanski (2014) suggest a general decline in regulation strategies from early to middle adolescence, with the smallest repertoire of emotion regulation strategies at the age of 15.

Although research has shown a general tendency for the quality of the emotional experience in early adolescence to deteriorate, the affective responses of adolescents also differ between contexts. School represents a major developmental context for adolescents (Eccles & Roeser, 2011), not least in terms of their emotional development (Moksnes et al., 2016). The landmark ESM study by Csikszentmihalyi and Larson (1984) already showed that adolescents were less emotionally positive in school compared with other contexts. More than 30 years later, a similar picture emerged in a study comparing adolescents' affective experiences in everyday life (Zurbriggen et al., 2018). As expected, adolescents experienced school time more negatively than leisure time.

Assessing Students' Emotional Experience in School

In contrast to many other scientific disciplines, research on emotions in education was slow to emerge, but has steadily increased in the last 20 years (Pekrun & Linnenbrink-Garcia, 2014). Although a number of studies have applied ESM to investigate adolescents' experiences in classrooms, other self-report instruments—primarily test anxiety questionnaires—have been much more widely used to measure students' emotions (Pekrun & Bühner, 2014). Moreover, situated multimethod approaches to the study of emotions in education are still relatively uncommon (Turner & Trucano, 2014). It is then hardly surprising that there are only a few studies comparing state reports (i.e., in situ reports) and general trait reports or retrospective reports of students' emotional experience in the classroom. Besides that, the few studies refer to different concepts of emotion in investigating the retrospection effects in students' emotion reports. Whereas the studies of Bieg et al. (2014) and Goetz et al. (2013) were based on the concept of discrete emotions, Venetz and Zurbriggen (2016) referred to the concept of core affect. It is assumed that an affective state is continually present, yet fluctuating across time, and consciously accessible as a simple, nonreflective feeling that can be described by basic bipolar dimensions (Russell, 2003; Yik et al., 1999, 2011).

The study of Goetz et al. (2013) revealed that girls, compared with boys, reported higher levels of trait anxiety, but not so in situ during mathematics instruction and testing situations. Furthermore, students' mathematics self-efficacy beliefs and self-concept were related to trait anxiety in mathematics. Bieg et al. (2014) reported that trait emotions of enjoyment, pride, anger, and anxiety are generally rated higher than state emotions, suggesting an intensity bias (see also Buehler & McFarland, 2001; Wirtz et al., 2003). As expected, the academic self-concept in mathematics predicted the discrepancy between trait and state reports of students' emotions. In contrast, Venetz and Zurbriggen (2016) found that students retrospectively report higher levels of positive affect and lower levels of negative affect compared with in situ ratings, pointing to a positive recall bias. The academic self-concept could predict the recall bias of positive affect, but only to a very small extent, while emotional attitudes toward school could predict the recall bias both in positive and negative affect.

The positive recall bias can also be referred to as a rosy retrospection or the rosy view effect (Mitchell & Thompson, 1994; Mitchell et al., 1997). According to Mitchell and Thompson's theory of temporal adjustment, individuals tend to remember, evaluate, and anticipate events (i.e., interrelated sequences of activities and behaviors) more positively than actually experienced. One central factor or boundary condition for such a rosy effect is personal involvement, because it "is fundamentally important for some of the motivational and cognitive processing that produces a rosy view" (Mitchell & Thompson, 1994, p. 89). Given the decline in students' motivation and emotional involvement in secondary education (Minnaert et al., 2017; Wigfield et al., 2006) and the increase in negative attitudes toward school (Hascher & Hadjar, 2018), we assume that the rosy view would decrease during early adolescence.

The Present Study

With the present study, we examined, first, whether students' retrospectively assessed affect in the classroom differed from their momentary, in situ reported affect (i.e., retrospection effect or recall bias). According to the theory of temporal adjustment (Mitchell & Thompson, 1994; Mitchell et al., 1997), we expected a positive bias (i.e., rosy view) for the retrospective affect ratings (Hypothesis 1). Such a rosy retrospection effect is characterized by higher positive affect and lower negative affect in retrospection than in situ. Moreover, we investigated whether the retrospection effect could be explained by students' emotional attitudes toward school (Hypothesis 2). Based on the accessibility model of Robinson and Clore (2002) and the theoretical ideas described above, we assumed that more positive attitudes toward school would be associated with greater retrospective overestimation of positive affect and more retrospective underestimation of negative affect.

Second, we examined whether the retrospection effect of students' affective experiences in the classroom changed during early adolescence. Given the developmental decline in the quality of emotional experience, we expected that retrospective overestimation of positive affect and underestimation of negative affect would become smaller during early adolescence (Hypothesis 3). Additionally, we investigated whether the change in retrospection effects could be predicted by a change in adolescents' emotional attitudes toward school in general. We assumed that change in attitudes toward school would be positively associated with the change in retrospection effects of positive affect (Hypothesis 4a). That is, we expected a general decline in the overestimation of positive affect, but this decline should be less pronounced for students with smaller declines (or even an increase) in positive attitudes toward school. Finally, we assumed that changes in attitudes toward school would be negatively associated with the change in retrospection effects of negative affect (Hypothesis 4b). That is, the underestimation of negative affect should generally decline over time, yet this decline should be less pronounced for students with smaller decreases (or even increases) in positive attitudes toward school.

Method

Sample and Procedure

The sample consisted of Swiss adolescents who participated in a study on emotional experience in the classroom and school-related self-perceptions (see Venetz et al., 2012). The project was funded by the Swiss National Research Foundation (Grant 122274). The sample was drawn from primary schools in rural and urban locations within four German-speaking states (cantons). The study was approved by the education department of the corresponding cantons (Grison, Thurgovia, St. Gallen, and Zurich) and by the school principals. Adolescents and their primary caregivers provided written informed consent.

The present analyses focus on data from 120 adolescents who participated twice within a period of 3 years. At the first occasion of measurement (T1), the 120 students (M_{age} = 12.1, SD_{age} = .89, 41.8% girls) were in Grade 5 (20.0%) or Grade 6 (80.0%), which in Switzerland corresponds to the two last years of primary education (Level 1 according to the *International Standard Classification of Education*, ISCED).[1] In the follow-up study (T2), after transition to lower secondary school (ISCED 2), the students were in Grades 8 or 9.

At both measurement occasions, the students participated in an experience sampling survey for 1 week and completed a conventional questionnaire at the end of the survey. At T1, the experience sampling took place during lessons on school days (Monday to Friday). On the last school day before the survey actually started, trained test administrators explained the basic idea and the procedure of the experience sampling, also in comparison with the conventional questionnaire. As part of the 45-min instruction in class, students completed a test signal that familiarized them with the ESM procedure. The 14 randomly generated signals of the experience sampling survey at T1 were sent via pager or a teacher's mobile phone (three signals per day, and on half-day lessons, only two). At the signal, teachers requested their students to complete a short questionnaire in a small booklet that students carried with them at all times during the five school days. It took 2 to 4 min to fill out one questionnaire. In total, the students completed N_{t1} = 1,488 experience sampling questionnaires (response rate T1: 88.6%).

At T2, the experience sampling additionally referred to leisure time. Trained test administrators instructed the adolescents individually 2 or 3 days before the start of the experience sampling survey. After signing a user agreement, adolescents were equipped with smartphones to report their emotional experience for seven consecutive days (including the weekend). At 42 randomly generated occasions (six per day), they received a link to a short online questionnaire via SMS. In total, the adolescents completed N_{t2} = 3930 experience sampling questionnaires in school and during leisure time (response rate T2: 78.0%). 78% of all questionnaires were completed within 10 min after receiving the link, 88% within 30 min. The 4.4% of the questionnaires that were answered more than 2 hr after receiving the link were not considered. To grant comparability between T1 and T2, only the n_{t2} = 1,083 occasions during lessons (response rate for lessons: 75.2%) were included for the present analyses.

Measures

Momentary Affect

Momentary affect was assessed using the Positive Activation (PA) and Negative Activation (NA) from the PANAVA short scales (PANAVA-KS; Schallberger, 2005). The scales are based on the two general activation systems of affect by Watson and colleagues (Watson & Tellegen, 1985; Watson et al., 1999).[2] The PA

[1] Class teachers who taught Grade 6 were asked to participate in the study of Venetz et al. (2012). Students from Grade 5 were only included if a class was a multigrade class (Grades 5 and 6). Multigrade classes are quite common in the rural parts of Switzerland.

[2] Schallberger (2005) included two items capturing valence (VA) in terms of pleasure as in the conceptualization by Russell (2003) in addition to eight items measuring PA and NA. In contrast to Russell's model assumptions, and in line with other studies (e.g., Schimmack & Grob, 2000), the evaluation of the PANAVA-KS indicated that the two VA items constituted a third (i.e., additional) dimension (Schallberger, 2005). To stay in line with previous research on intraindivdual variation in emotional experience (e.g., Lay et al., 2017; Riediger et al., 2009; Röcke et al., 2011) and because the validity of the VA scale is not yet fully clear, we focus on PA and NA as in the model by Watson and Tellegen (1985).

scale consists of four items with pairs of opposite adjectives (e.g., *listless* vs. *highly motivated, bored* vs. *excited*; see Table 1). The NA scale also comprises four items (e.g., *relaxed* vs. *stressed*, *calm* vs. *nervous*). All eight items were scored on a 7-point bipolar Likert scale ranging from −3 to 3, with 0 labeled as *neither nor*. For the current analyses, the items have been transformed to positive values ranging from 1 to 7. As an introductory question, adolescents were asked: *How did you feel just before the signal?*

At T1, reliability coefficients (McDonald's ω) for the PA scale ranged from ω = .66 (95% confidence interval, CI [.54, .78]) to ω = .84 (95% CI [.78, .90]) across the 14 randomly generated signals. For the NA scale this range was ω = .60 (95% CI [.46, .75]) to ω = .80 (95% CI [.71, .90]). At T2 reliability coefficients for the PA scale ranged from ω = .79 (95% CI [.71, .88]) to ω = .87 (95% CI [.83, .91]). For the NA scale this range was ω = .68 (95% CI [.58, .78]) to ω = .85 (95% CI [.78, .91]). We report the range of McDonald's ω (McDonald, 1999) for the in situ measurements to reflect the fluctuation of reliabilities across the signals. The "average" in situ measurements for every student are modeled as error free variables with perfect reliability at Level-2 in multilevel models (see also Eid et al., 2008; Koch et al., 2014).

Retrospective Affect

At the end of both experience sampling phases, adolescents retrospectively reported their affective experience in class by means of a conventional questionnaire. The same PA and NA items were used as for the momentary ratings. Adolescents were asked to rate how they experienced lessons in general during the last week. In the introductory text to this question and in the instruction by the test administrators (see also section "Sample and Procedure"), it was stressed that these ratings should refer to the week when the experience sampling study was held.

The reliability for the PA scale was ω = .78 (95% CI [.70, .86]) at T1 and ω = .80 (95% CI [.73, .86]) at T2. The reliability for the NA scale was ω = .74 (95% CI [.63, .84]) at T1 and ω = .51 (95% CI [.37, .66]) at T2.

Emotional Attitudes toward School

Emotional attitudes toward school were also measured at the end of both experience sampling phases using a scale from the Perceptions of Inclusion Questionnaire (PIQ; Venetz et al., 2015). The scale refers to the (explicit) affective component of attitudes and consists of four items assessing emotional well-being in school in general (e.g., *I like going to school* or *School is fun*; see Table 1 for item wording). For each item, participants were asked to provide a rating on a 4-point rating scale (0 = *not at all true*, 1 = *somewhat not true*, 2 = *somewhat true*, 3 = *certainly true*). For the current analyses, the items have been transformed to positive values ranging from 1 to 4. Studies have reported good reliability and validity for the PIQ (DeVries et al., 2018; Zurbriggen et al., 2019). In the current study, the reliability was ω = .92 (95% CI [.90, .95]) at T1, and ω = .88 (95% CI [.83, .94]) at T2.

Analyses

To test our hypotheses, we estimated three consecutive multilevel structural equation models. We first tested for differences in the factorial structure across the two occasions of measurement (baseline model, see Figure 1) estimating multilevel confirmatory factor analytic (CFA) models for the two occasions of measurement as a baseline model. Consider T1: On Level 1, we specified the *i*n *s*itu *a*ssessments of *p*ositive *a*ctivation (isapa11–isapa41) and *n*egative *a*ctivation (isana11–isana41). The four indicators load on their respective factor (L1PAT1 or L1NAT1). These factors depict the in situ deviation of PA and NA from the respective Level-2 factor (L2PAT1 or L2NAT1; i.e., the aggregated—or "average"— momentary affect across the in situ measurements). Also on Level 2, the four indicators of the retrospectively assessed PA (rpa11–rpa41: *r*etrospective *p*ositive

Table 1

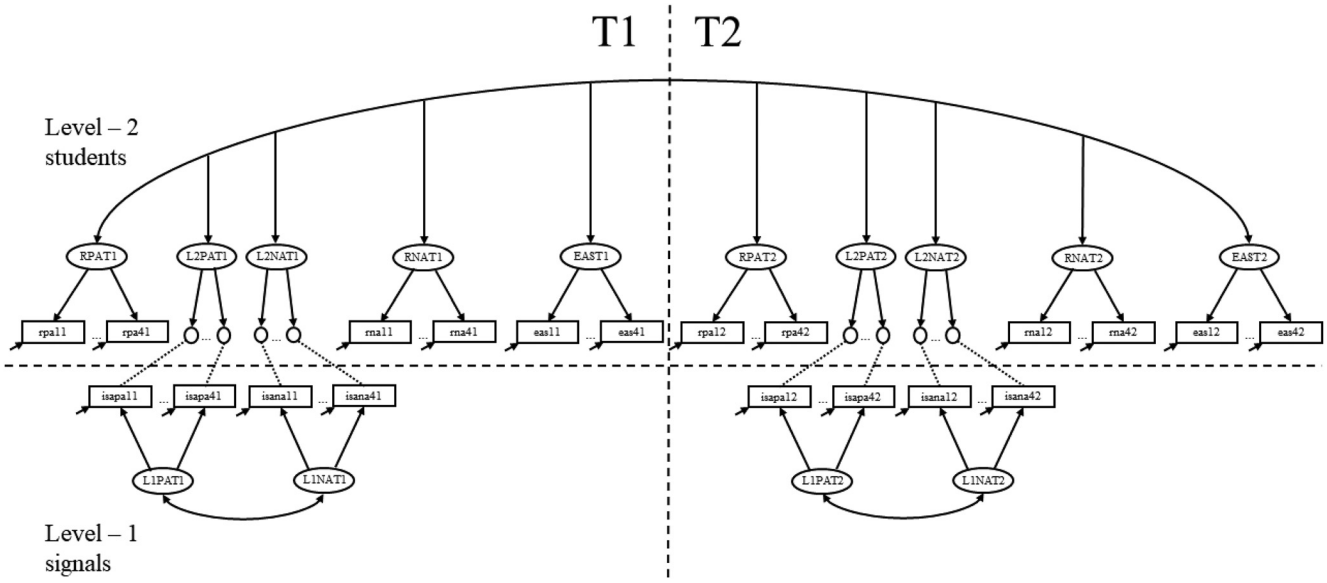
Unstandardized Loadings and Intercepts of Study Variables in the Longitudinal Multilevel CFA (Baseline Model), and ICC Coefficients for the In Situ Measurements of Positive and Negative Activation

Variable	Loading	Intercept	ICC at T1	ICC at T2
PA1 (energetic–weakened)	1	4.884	.206	.242
PA2 (tired–wide awake)[a]	1.078	4.336	.377	.217
PA3 (listless–highly motivated)[a]	1.182	4.583	.272	.248
PA4 (excited–bored)	1.065	4.267	.237	.190
NA1 (stressed–relaxed)	1	2.900	.220	.160
NA2 (peaceful–angry)[a]	1.009	2.526	.249	.201
NA3 (calm–nervous)[a]	0.717	2.704	.263	.205
NA4 (worried–carefree)	0.868	2.791	.287	.293
EAS1 (I like going to school)	1	2.884		
EAS2 (I have no desire to go to school)[a]	1.116	2.977		
EAS3 (I like it in school)	1.093	2.947		
EAS4 (school is fun)	1.079	2.562		

Note. N = 120. CFA = confirmatory factor analysis; PA = Positive Activation; NA = Negative Activation; ICC = intraclass correlation. Note that the original items are in German and that the translations are not validated. The first loading parameter were fixed to unity (1) for each latent variable for model identification. All estimated coefficients were significantly different from zero (all $p < .001$). For the PA- and NA-items strong measurement invariance was implemented across measurement methods (in situ vs. retrospective), levels (Level 1 vs. Level 2; see Figure 1) and measurement occasions (T1 vs. T2). For the EAS-items strong measurement invariance was implemented across measurement occasions. Hence, PA1 to PA4 and NA1 to NA4 refer to all ispa, rpa, isna, and rna-items, respectively; EAS1 to EAS4 refer to all EAS items.

[a] Reversed item.

Figure 1
The Baseline Model



Note. rpa11 = the first indicator for retrospective positive activation at T1; rpa41 = the fourth indicator for retrospective positive activation at T1; isapa11 = the first indicator for the in situ assessment of positive activation at T1; isapa41 = the fourth indicator for the in situ assessment of positive activation at T1; isana11 = first indicator of the in situ assessment of negative activation at T1; isana41 = the fourth indicator for the in situ assessment of negative activation at T1; rna11 = the first indicator for retrospective negative activation at T1; rna41 = the fourth indicator for retrospective negative activation at T1; eas11 = the first indicator for emotional attitudes towards school at T1; eas41 = the fourth indicator for emotional attitudes towards school at T1; RPAT1 = retrospective positive activation at T1; L1PAT1 = Level 1 positive activation at T1; L1NAT1 = Level 1 negative activation at T1; L2PAT1 = Level 2 positive activation at T1; L2NAT1 = Level 2 negative activation at T1; RNAT1 = retrospective negative activation at T1; EAST1 = emotional attitudes towards school at T1. The same measurement model is repeated for T2. The circles under the Level–2 latent variables L2PAT1, L2NAT1, L2PAT2, and L2NAT2 refer to the "aggregated" in situ assessments across the Level–1 units (see also Eid et al., 2008). The three dots (. . .) represent the second and the third indicator of a respective measure, which are not depicted. The term "signals" refers to the in situ assessments. The third level (classes) is not depicted, because in the replicate weights method (Asparouhov & Muthén, 2010) the Level-3 clustering is statistically respected but not explicitly modeled.

*a*ctivation) or NA (rna11–rna41: *r*etrospective *n*egative *a*ctivation) load on their corresponding factor (RPAT1 or RNAT1; i.e., retrospective affect). Emotional attitudes toward school (EAS) are also modeled at Level 2 (EAST1 with its indicators eas11–eas41). For T2, exactly the same models were specified (with e.g., L2PAT2 representing the Level-2 factor for in situ PA at T2).

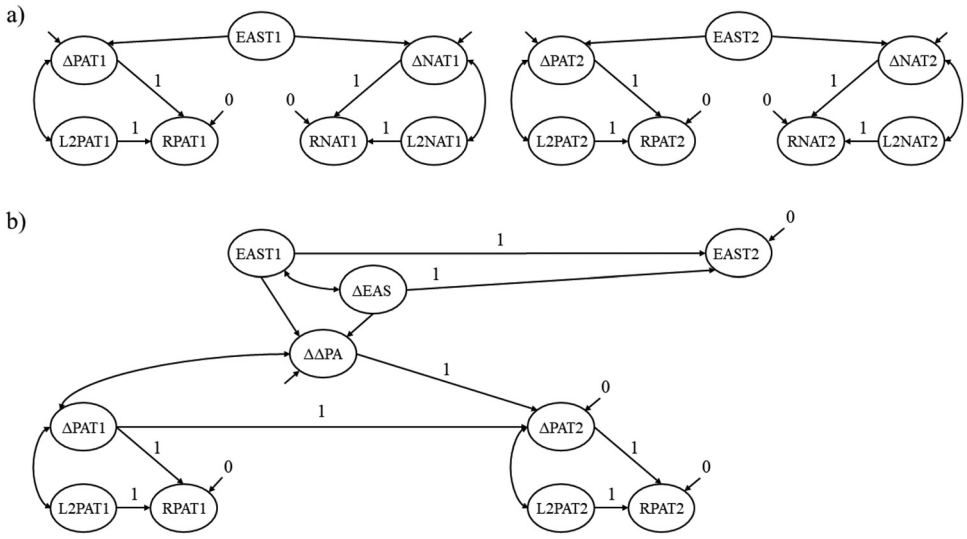
In the second model (first-order difference model, see Figure 2a), we introduced the latent difference variables (e.g., ΔPAT1 and ΔNAT1 for T1) at Level 2 depicting the latent difference of the retrospectively reported PA and NA (e.g., RPAT1 or RNAT1) from the average in situ positive and negative affect (e.g., L2PAT1 or L2NAT1) for the two occasions of measurement separately (McArdle & Nesselroade, 1994). These differences directly represent the retrospection effects at T1 or T2 (Hypothesis 1). According to Hypothesis 2, the four differences are to be regressed on the EAS at the respective point in time.

In the third model (second-order difference model, see Figure 2b), we included differences between the latent variables of T2 and T1. That is, ΔEAS represents the change in the emotional attitudes toward school, and ΔΔPA and ΔΔNA represent the change in retrospection effects across the two occasions of measurement (Hypothesis 3). The two differences in affect across time (ΔΔPA and ΔΔNA) are regressed on the initial level and the difference of emotional attitudes toward school across time (Hypotheses 4a and 4b).

To estimate latent differences, interpret mean values and compare correlations across time, assumptions of strong measurement invariance must hold (Geiser et al., 2014; Meredith, 1993). These assumptions were tested in the first model (baseline model; see online supplemental material). We assessed goodness of fit with the χ^2 value as well as the root mean square error of approximation (RMSEA) and the gamma hat-index ($\hat{\gamma}$). The latter was included because it has been shown that $\hat{\gamma}$ is particularly sensitive toward model miss-specifications but insensitive toward fit-irrelevant factors like sample size and model type (Fan & Sivo, 2007). Benchmark-values for the fit indices were taken from West et al. (2012), Schermelleh-Engel et al. (2003), Fan & Sivo (2007), and Marsh et al. (2004). RMSEA $< .08$ and $\hat{\gamma} > .90$ indicate acceptable model fit and RMSEA $< .05$ and $\hat{\gamma} > .95$ indicate good model fit. All models were specified in M*plus* Version 8.2 (Muthén & Muthén, 1998–2017; model syntaxes are presented in the online supplemental material). The data can be obtained from the first author by request.

As several students attended the same classes at T1, they were also nested within classes yielding a three-level data structure. We respected the nested data structure at T1 (in all models) using the replicate weights method (COMPLEX-option; Asparouhov &

Figure 2
(a) First-Order Difference Model and (b) Second-Order Difference Model



Note. For the sake of simplicity, no item has been depicted and only the latent structure is displayed, yet, exactly the same measurement structure as for the baseline model is estimated. Latent variables with the same names as in the baseline model (Figure 1) correspond to the same variables of the baseline model. ΔPAT1 = change in retrospective positive activation in comparison with in situ positive activation at T1; ΔNAT1 = change in retrospective negative activation in comparison with in situ negative activation at T1; ΔPAT2 = change in retrospective positive activation in comparison with in situ positive activation at T2; ΔNAT2 = change in retrospective negative activation in comparison with in situ negative activation at T2; ΔEAS = change in emotional attitudes towards school at T2 compared with T1. ΔΔPA = change in the retrospection effect for positive activation at T2 compared with T1. The second-order difference model for NA (not depicted) is exactly the same as for PA. Note that in 2a, residuals of the difference variables may correlate (not depicted).

Muthén, 2010). At T2 very few students attended the same classes preventing us from respecting the (possibly differing) nestedness at T2. We present the results of an analysis with only two levels (in situ assessments in students) in the online supplemental material. No substantial differences can be found.

Results

Model Fit

Table 2 presents fit indices for the baseline model (with final measurement invariance restrictions), the first-order difference model and the second-order difference model. While the test of exact fit (χ^2) indicates misfit, RMSEA and $\hat{\gamma}$ suggest satisfactory fit for the three models.

Baseline Model

In Table 1, we present the unstandardized loading parameters and intercepts reflecting strong measurement invariance (a) for emotional attitudes toward school across time (T1 and T2), and (b) for PA and NA across time (T1 and T2) *and* methods (in situ assessment and retrospective assessment). Furthermore, the last two columns of Table 1 depict intraclass correlation (ICC) coefficients for in situ measurements of positive and negative activation at T1 and T2, respectively. Between 16.0% and 37.7% of the variation is attributed to differences between individuals (L2). While these proportions are substantial, the answers to these items are mostly influenced by situational factors.

In Table 3, we present latent means, correlations and variances from the baseline model. In the last line, the latent means reveal that students score above the scale midpoint (4) on all measures

Table 2
Fit Evaluation for the Baseline Model, the First-Order Difference Model, and the Second-Order Difference Model

Model	$\chi^2(df)$	p	RMSEA	$\hat{\gamma}$
Baseline model	3,122.668 (874)	<.001	.039	.938
First-order difference	3,103.345 (870)	<.001	.039	.938
Second-order difference	3,102.850 (872)	<.001	.039	.938

Note. N = 120. RMSEA = root mean square of error approximation.

Table 3
Variances (in the Diagonal), Correlations, and Means (in the Last Row) of Latent Variables in the Longitudinal Multilevel CFA (Baseline Model)

Variable	L1PAT1	L1NAT1	L1PAT2	L1NAT2	L2PAT1	RPAT1	L2NAT1	RNAT1	EAST1	L2PAT2	RPAT2	L2NAT2	RNAT2	EAST2
L1PAT1	0.757	−.865***												
L1NAT1		0.884												
L1PAT2			0.998	−.602***										
L1NAT2				1.029										
L2PAT1					0.716	.897***	−.530***	−.410**	.573***	.128	.058	.161	.243	.303*
RPAT1						1.139	−.552***	−.637***	.730***	.122	.063	.072	.064	.274
L2NAT1							0.670	.866***	−.379**	−.030	.170	.037	−.081	−.092
RNAT1								1.204	−.404***	−.103	.033	.158	.057	−.142
EAST1									0.569	.100	.325**	−.099	−.063	.467***
L2PAT2										0.609	.368**	−.784***	−.263*	−.045
RPAT2											0.836	−.339*	−.692***	.695***
L2NAT2												0.530	.314*	.047
RNAT2													0.539	−.279**
EAST2														0.324
M					5.262	5.378	2.756	2.705	3.027	4.584	4.712	3.113	3.167	2.976

Note. $N = 120$. CFA = confirmatory factor analysis. The diagonal displays latent variable variances (all $p < .001$). The first four latent variables are on the within-level the remaining variables are on the between-level. Latent variables from different levels do not correlate with each other.
* $p < .05$. ** $p < .01$. *** $p < .001$.

of PA and lower on all measures of NA. In the upper left part, the correlation coefficients of the Level-1 latent variables show that within one occasion of measurement (in situ assessment), NA and PA were strongly inversely related ($r = -.865$ at T1 and $r = -.602$ at T2). Since the Level-1 latent variables reflect random fluctuations they are not allowed to covary across occasions of measurement.

Latent correlations at Level 2 indicate that the retrospective and aggregated in situ assessments were strongly associated at T1 ($r = .897$ for PA and $r = .866$ for NA) yet to a much smaller extent at T2 ($r = .368$ for PA and $r = .314$ for NA). The correlations between variables representing PA and those representing NA at T1 or at T2, respectively, were all negative. In particular, we found strong negative correlations within the assessment methods—that is, aggregated in situ variables correlated strongly with each other as retrospective assessments did. Additionally, at T1 we found a strong negative correlation between aggregated in situ assessed NA and retrospectively assessed PA. Furthermore, we found that the Level-1 variances increased over the two occasions of measurement whereas Level-2 variances decreased indicating that the in situ assessments at T2 became more volatile compared with T1.

The correlations of emotional attitudes toward school with the other latent variables show an interesting pattern. At T1, attitudes toward school correlated positively with the two latent variables representing PA and negatively with those representing NA (all $r > .379$). At T2, attitudes only correlated with the retrospective assessments but no longer with the aggregated in situ assessments.

First-Order Difference Model

The first-order difference model is particularly informative with regard to the retrospection effects at the two occasions of measurement. At T1, the retrospectively reported PA was on average .207 scale units higher than the aggregated in situ reported PA ($p = .047$; $d = .412$). For NA, we found that the average retrospectively reported score was .154 scale units below the aggregated in situ reported score ($p = .131$, $d = -.269$), yet this difference was not statistically significant. At T2, the latent differences descriptively changed directions. The retrospectively reported PA was on average .220 scale units lower than the aggregated in situ reported PA ($d = -.220$). However this mean difference did not reach significance ($p = .058$). For NA, the average retrospectively reported score was significantly higher than the aggregated in situ score ($M(\Delta\text{na}) = .244$, $p = .032$, $d = .282$).

Predicting the retrospective bias, we found that emotional attitudes toward school predicted the retrospective bias for PA ($b = .336$, $\beta = .520$, $p = .002$) but not for NA ($b = -.134$, $\beta = -.177$, $p = .273$) at T1. At T2, we found the same pattern of regression coefficients with both coefficients being statistically significant (Δpa: $b = 1.242$, $\beta = .703$, $p < .001$; Δna: $b = -.492$, $\beta = -.322$, $p = .024$).

Second-Order Difference Model

As already reported, the latent difference variables depicting retrospection effects in the first-order difference models changed directions at T2. The second-order difference model allows for a statistical test of the change in retrospective bias via the second-order difference variables. We found statistically significant

changes in retrospective effects both for PA ($M(\Delta\Delta pa) = -.233$, $p = .031$, $d = -.212$) and NA ($M(\Delta\Delta na) = .274$, $p = .018$, $d = .264$).

Predicting the changes in the retrospective effects, we found that students with initially more positive emotional attitudes toward school showed a smaller decrease in their retrospective bias for PA ($b = .986$, $\beta = .680$, $p < .001$). Yet, emotional attitudes toward school did not predict change in retrospective bias for NA ($b = -.271$, $\beta = -.197$, $p = .418$). Moreover, we found that increases in positive emotional attitudes toward school also led to a smaller decrease in the retrospective bias for PA ($b = 1.561$, $\beta = .989$, $p < .001$) but did not predict changes in the retrospective bias for NA ($b = -.590$; $\beta = -.394$, $p = .058$). Changes in the positive and negative retrospective bias correlated more strongly with each other than predicted by the initial attitudes and changes in emotional attitudes toward school (partial correlation: $r = -.670$, $p < .001$).

Discussion

The present study is the first, to our knowledge, to investigate change in retrospection effects or recall bias of emotional self-reports during early adolescence. To this end, we focused on school as a major context in early adolescence. We compared the retrospectively assessed positive and negative affect of 120 students with their momentary, in situ reported positive and negative affect in the classroom at two measurement occasions in early adolescence, or to be more precise, at the age of about 12 years (T1) and 3 years later at the age of about 15 years (T2). Furthermore, we examined whether students' emotional attitudes toward school in general have an effect on recall bias of students' affect ratings as well as on *change* in recall bias during early adolescence. To test our hypotheses, we applied first-order and second-order latent difference models in a multilevel structural equation modeling framework.

Changes in Recall Bias of PA and NA in Early Adolescence

Based on the results of the baseline model, we concluded that strong measurement invariance was tenable for emotional attitudes toward school across time (i.e., T1 and T2) and for PA and NA also across methods (i.e., in situ and retrospective assessment; see online supplemental material). An interesting secondary finding of the baseline model is that the in situ assessments were more volatile at T2 than at T1, that is, adolescents' affective experiences fluctuated more across situations at the end of early adolescence than at the beginning. The relatively high affective variability has often been described as a central feature of adolescence (e.g., Csikszentmihalyi & Larson, 1984; Maciejewski et al., 2015; McLaughlin et al., 2015; Weinstein et al., 2007). The higher intra-individual variability in students' affective experiences could be associated with a decline in the repertoire of emotion regulation strategies from early to middle adolescence (Zimmermann & Iwanski, 2014).

As expected, results of our study point toward a positive recall bias at the beginning of early adolescence (see also Venetz & Zurbriggen, 2016). At the age of about 12 years, students reported higher PA retrospectively at the end of one school week than in situ during class. At the same time, their NA was lower when reported retrospectively than in situ, but this difference was not statistically significant. Having said that, results (of T1) can be interpreted as a small positive recall bias in terms of a rosy view (Mitchell & Thompson, 1994; Mitchell et al., 1997). Three years later (at T2), retrospection effects changed direction: Students at the age of about 15 years reported lower PA (though not significant) and higher NA retrospectively than in situ. Hence, in contrast to T1, at T2 we found evidence for a small negative recall bias in terms of a "blue retrospection effect." Taken together, these results indicate that, in early adolescence, positive affect seems to be overestimated whereas in middle adolescence negative affect seems to be overestimated.

Although these results obtained by the first-order difference model provide some insight into the recall bias of students' affect ratings at the beginning and at the end of early adolescence, they do not give indications for a possible *change* in recall bias during early adolescence. Therefore, we also applied a second-order difference model. The findings revealed effects with reversed indications: On average across students, we found a decrease in the positive retrospection effect from T1 to T2 (negative change), but an increase in the negative retrospection effect (positive change). The small effect sizes point to a negative shift in recall bias—or to a "blue view"—of students' affect ratings during early adolescence. This may be related to the general decline in the positivity of emotional experiences in adolescence (Hektner, 2014; Larson et al., 2002), or to a preference for dampening positive affect and for maintaining or enhancing negative affect. Previous research demonstrated that such a contra-hedonic orientation is relatively prevalent in adolescence compared with older age groups (Riediger et al., 2009). Adolescents may seek to enhance negative affect or dampen positive affect because it is socially appropriate. Contra-hedonic orientation may also help adolescents with low self-esteem to achieve consistency in their emotional self-perception (Riediger et al., 2011; Wood et al., 2009).

The change in recall bias, reflected as an increase in the negative retrospection effect, could also be associated with a decrease in affective control, since the findings of a recent review by Schweizer et al. (2020) suggest that affective control, most notably inhibition of affective information, is reduced during adolescence. Furthermore, it might be worthwhile to investigate whether emotional granularity (often used synonymously with emotional differentiation) is linked with *changes* in retrospection effects in adolescence. Both strong affective control and strong emotion granularity or differentiation, especially differentiation of negative emotions (Nook et al., 2018), are of particular importance for adolescents' mental health and well-being (Lennarz et al., 2018; Tugade et al., 2004). Despite the widely recognized importance of understanding emotion regulation in adolescence, we still lack knowledge about how exactly the use of emotion regulation and coping skills develop from middle childhood to adolescence (Compas et al., 2017; Lennarz et al., 2019).

Emotional Attitudes Toward School as a Predictor of Recall Bias

Investigating recall bias of emotional self-reports, we focused on school as one major developmental context for adolescents. Since school tends to become emotionally negative during early

adolescence, while leisure time and peer contexts beyond school gain importance, it can be assumed that a positive recall bias (i.e., rosy view) is less likely to occur at the end of early adolescence. According to Mitchell and Thompson's (1994) theory of temporal adjustment, personal involvement is fundamental for the "rosy phenomenon." Results of our study support this hypothesis. The rosy view at the age of about 12 could be, to some extent, explained by students' attitudes toward school *in general*, that is, their inner tendency to evaluate school with some degree of (dis-) favor as defined by Eagly and Chaiken (2007). Three years later, adolescents' emotional attitudes toward school again had a large, positive effect on their retrospective assessment of PA, but also a moderate, negative effect on their retrospective assessment of NA. Moreover, students' emotional attitudes toward school were correlated only with the retrospective assessment of affective experience, but no longer with the in situ assessment (as at T1). Taken together, these findings indicate that at the age of 15, students rely more on attitudes toward school to judge their emotional experience in class than they do at the age of 12.

Having said that, positive emotional attitudes toward school seem to serve as buffer for the "blue view" at the end of early adolescence. Results revealed that students with initially more positive emotional attitudes toward school showed a smaller decrease in their recall bias of PA across early adolescence. Moreover, an increase in attitudes was associated with a smaller decrease in their recall bias of PA. In line with these findings, positive attitudes toward school could be considered to have a preventive function against the general decline in emotional positivity in early adolescence.

Limitations and Future Directions

The current study has several limitations. First, the mode of data collection was different at T1 and T2. At T1, students reported their positive and negative affect in situ and retrospectively as well as their emotional attitudes toward school on paper-and-pencil questionnaires, whereas at T2 the constructs were assessed online. One could also argue that both experience sampling phases are not comparable, because the experience sampling at T2 also covered leisure time, even though only the signals received during lessons were included in our analyses. It needs to be stressed that results demonstrated that emotional attitudes toward school were measurement invariant across time (i.e., T1 and T2) and PA and NA also across methods (i.e., in situ and retrospective assessment).

Second, we referred to the concept of core affect to assess adolescents' affective experience. While the concept of core affect is well suited to capture affective experiences in any situation of everyday life (Kuppens et al., 2012), comparisons of our findings with those from previous research are limited. The few other studies on recall bias of students' emotional experiences, that is, differences pertaining to students' state and trait emotion reports, were (cross-sectional and) based on the concept of discrete emotions assessing specific academic emotions in mathematics (Bieg et al., 2014; Goetz et al., 2013). Besides that, they relied on trait reports of emotions that students *generally* felt during (mathematics) lessons and not on retrospective reports referring to a specific time frame (e.g., one school week). Future studies on retrospection effects might benefit from combining both concepts, that is, to assess discrete emotions as well as PA and NA (i.e., core affect). At this point, it should be noted that the reliability of the NA scale assessed retrospectively at T2 was rather low as were some of the in situ reliabilities. This could hint toward a differentiation of (recalled) negative affect into a more complex and multidimensional construct (i.e., reflecting several discrete emotions) during adolescence. Although the relation of the indicators to the latent variable remains the same across time (measurement invariance), the amount of error variance increases reflecting the heterogeneity of the items at T2 (i.e., complexity). Hence, more research is needed to investigate if negative affect is a one-dimensional construct in adolescents, and if retrospection effects vary between the different aspects of PA and NA. Due to the fact, that these aspects were measured by single items, we could not properly investigate these potentially differential retrospection effects relying on latent variables. Yet, we provide results of manifest change models at the item-level in the online supplemental material. Further research could also address questions related to the two items capturing valence (i.e., pleasure), the third dimension of the PANAVA short scales (Schallberger, 2005), or questions regarding (changes in) the relation between PA and NA (see, e.g., Dejonckheere et al., 2021).

Third, the focus on the school context does not permit us to generalize our findings to other central contexts of an adolescent's daily life. Among others, the question remains whether the positive recall bias or rosy view at the beginning of early adolescence also occurs in leisure time. Moreover, it could be argued that the negative shift in affect ratings is largely due to the transition from primary to secondary education. Findings from an international comparative study by Casas and González-Carrasco (2019) suggest, however, that the general decline in well-being from childhood to adolescence is not related to school transition. Nonetheless, negative emotions such as test anxiety or worries about the future are more prevalent in lower secondary school than in primary school, particularly in transition to further education or to work life. As a similar "negativity bias" has often be observed among young adults (e.g., Charles et al., 2016; Neubauer et al., 2020), future longitudinal studies might investigate whether the negative tendency of emotional self-reports persists in late adolescence.

Fourth, we only accounted for emotional attitudes toward school as a predictor of recall bias of students' affect ratings. Choosing this predictor was based on the theoretical considerations that bias of retrospective reports is due to semantic retrieval strategies, and that school is associated with a particular set of beliefs (i.e., situation-specific beliefs; Robinson & Clore, 2002). Since we measured *positive* attitudes toward school (e.g., "I like going to school"), one might argue that emotional attitudes and PA both reflect a broader positive view or trait that may in turn explain the higher correlations of attitudes toward school with PA compared with NA (see Table 3). Furthermore, attitudes toward school is context-specific, hence, future studies should also investigate more general (context-free) influences on emotional attitudes, such as identity-related beliefs. If a situation or context is not associated with particular beliefs, a person is more likely to rely on identity-related beliefs when judging emotions or affective experiences in retrospection. Correspondingly, previous studies on recall bias have often examined self-esteem (e.g., Christensen et al., 2003) or personality trait associations (e.g., Lay et al., 2017; Mill et al., 2015). Future studies could simultaneously examine the impact of situation-specific and identity-related beliefs. In a similar vein, future studies might also consider accounting for peak and recency effects during the recalled period to explain bias of

retrospective affect ratings (Fredrickson, 2000). According to Robinson and Clore (2002), peak and recency effects are increasingly less likely to exert an influence on retrospective affect ratings when episodic details are relatively inaccessible. Therefore, studies with shorter intervals or event-contingent sampling are required. In the school context, high intense affective states experienced on painful events (e.g., being bullied) or in particularly demanding situations (e.g., exams) could presumably be a source of recall bias.

Finally, it would be worthwhile to consider time-varying predictors to investigate within-person fluctuations in adolescents' retrospective affect ratings or to examine within-person correspondence convergence. To investigate time-varying (event-level) predictors, relatively narrow and discontinuous (instead of time-inclusive and aggregated) retrospective reports are to be used. In this case, recall bias corresponds to the discrepancy between the momentary affect rating and the retrospective rating referring to this single momentary rating (Ottenstein & Lischetzke, 2020). Potential time-varying predictors would be the personal relevance of an event (Ottenstein & Lischetzke, 2020) or response time, as a proxy for emotional clarity (Arndt et al., 2018). Future research might also consider examining the within-person correspondence convergence of momentary reports and *several* retrospective end-of-week or end-of-day reports (Neubauer et al., 2020). For example, it remains an open question whether momentary and retrospective affect ratings share a substantial amount of within-person variance, or to put it in another way, whether both assessment types are suited to differentiate between "good weeks" from "bad weeks."

Conclusion

The current study extends previous research on recall bias by investigating change in retrospection effects of students' affective experience during early adolescence. To examine retrospection effects and change in retrospection effects, we applied first-order and second-order latent difference models in a multilevel structural equation modeling framework. Our findings point to a positive recall bias (i.e., rosy view) of students' affect ratings at the beginning of early adolescence and a negative shift in recall bias by the end of early adolescence. In other words, recall bias changed from a rosy view to a blue view. Furthermore, our findings support the hypothesis that a rosy view is less likely to occur, if a student is less emotionally involved in school. In turn, positive emotional attitudes toward school seem to serve as a buffer against the tendency toward a negative recall bias.

References

Arndt, C., Lischetzke, T., Crayen, C., & Eid, M. (2018). The assessment of emotional clarity via response times to emotion items: Shedding light on the response process and its relation to emotion regulation strategies. *Cognition and Emotion*, *32*(3), 530–548. https://doi.org/10.1080/02699931.2017.1322039

Asparouhov, T., & Muthén, B. (2010). *Resampling methods in Mplus for complex survey data*. Muthén & Muthén. Retrieved from https://www.statmodel.com/download/Resampling_Methods5.pdf

Bertha, E. A., & Balázs, J. (2013). Subthreshold depression in adolescence: A systematic review. *European Child & Adolescent Psychiatry*, *22*(10), 589–603. https://doi.org/10.1007/s00787-013-0411-0

Bieg, M., Goetz, T., & Lipnevich, A. A. (2014). What students think they feel differs from what they really feel—Academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PLoS ONE*, *9*(3), e92563. https://doi.org/10.1371/journal.pone.0092563

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods. An introduction to diary and experience sampling research*. Guilford Press.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin*, *27*(11), 1480–1493. https://doi.org/10.1177/01461672012711009

Casas, F., & González-Carrasco, M. (2019). Subjective well-being decreasing with age: New research on children over 8. *Child Development*, *90*(2), 375–394. https://doi.org/10.1111/cdev.13133

Charles, S. T., Piazza, J. R., Mogle, J. A., Urban, E. J., Sliwinski, M. J., & Almeida, D. M. (2016). Age differences in emotional well-being vary by temporal recall. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *71*(5), 798–807. https://doi.org/10.1093/geronb/gbv011

Christensen, T. C., Wood, J. V., & Barrett, L. F. (2003). Remembering everyday experience through the prism of self-esteem. *Personality and Social Psychology Bulletin*, *29*(1), 51–62. https://doi.org/10.1177/0146167202238371

Compas, B. E., Jaser, S. S., Bettis, A. H., Watson, K. H., Gruhn, M. A., Dunbar, J. P., Williams, E., & Thigpen, J. C. (2017). Coping, emotion regulation, and psychopathology in childhood and adolescence: A meta-analysis and narrative review. *Psychological Bulletin*, *143*(9), 939–991. https://doi.org/10.1037/bul0000110

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*, 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: methods for studying everyday life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. Wiley. https://doi.org/10.1002/9781118900772.etrds0010

Csikszentmihalyi, M., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. Basic Books.

Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., & Kalokerinos, E. K. (2021). The relation between positive and negative affect becomes more negative in response to personally relevant events. *Emotion*, *21*(2), 326–336. https://doi.org/10.1037/emo0000697

DeVries, J. M., Voß, S., & Gebhardt, M. (2018). Do learners with special education needs really feel included? Evidence from the Perception of Inclusion Questionnaire and Strengths and Difficulties Questionnaire. *Research in Developmental Disabilities*, *83*, 28–36. https://doi.org/10.1016/j.ridd.2018.07.007

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602. https://doi.org/10.1521/soco.2007.25.5.582

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, *21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*(3), 230–253. https://doi.org/10.1037/a0013219

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808

Geiser, C., Burns, G. L., & Servera, M. (2014). Testing for measurement invariance and latent mean differences across methods: Interesting incremental information from multitrait-multimethod studies. *Frontiers in Psychology*, *5*, 1216. https://doi.org/10.3389/fpsyg.2014.01216

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Goldbeck, L., Schmitz, T. G., Besier, T., Herschbach, P., & Henrich, G. (2007). Life satisfaction decreases during adolescence. *Quality of Life Research*, *16*, 969–979. https://doi.org/10.1007/s11136-007-9205-5

Hascher, T., & Hadjar, A. (2018). School alienation—Theoretical approaches and educational research. *Educational Research*, *60*(2), 171–188. https://doi.org/10.1080/00131881.2018.1443021

Hektner, J. M. (2014). Developmental psychology. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 585–600). Guilford Press.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method. Measuring the quality of everyday life*. Sage. https://doi.org/10.4135/9781412984201

Holsen, I., Kraft, P., & Vittersø, J. (2000). Stability in depressed mood in adolescence: Results from a 6-year longitudinal panel study. *Journal of Youth and Adolescence*, *29*(1), 61–78. https://doi.org/10.1023/A:1005121121721

Koch, T., Schultze, M., Eid, M., & Geiser, C. (2014). A longitudinal multilevel CFA-MTMM model for interchangeable and structurally different methods. *Frontiers in Psychology*, *5*, 311. https://doi.org/10.3389/fpsyg.2014.00311

Kuppens, P., Champagne, D., & Tuerlinckx, F. (2012). The dynamic interplay between appraisal and core affect in daily life. *Frontiers in Psychology*, *3*, 380. https://doi.org/10.3389/fpsyg.2012.00380

Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, *73*(4), 1151–1165. https://doi.org/10.1111/1467-8624.00464

Larson, R., & Ham, M. (1993). Stress and "storm and stress" in early adolescence: The relationship of negative events with dysphoric affect. *Developmental Psychology*, *29*(1), 130–140. https://doi.org/10.1037/0012-1649.29.1.130

Lay, J. C., Gerstorf, D., Scott, S. B., Pauly, T., & Hoppmann, C. A. (2017). Neuroticism and extraversion magnify discrepancies between retrospective and concurrent affect reports. *Journal of Personality*, *85*(6), 817–829. https://doi.org/10.1111/jopy.12290

Lennarz, H. K., Hollenstein, T., Lichtwarck-Aschoff, A., Kuntsche, E., & Granic, I. (2019). Emotion regulation in action: Use, selection, and success of emotion regulation in adolescents' daily lives. *International Journal of Behavioral Development*, *43*(1), 1–11. https://doi.org/10.1177/0165025418755540

Lennarz, H. K., Lichtwarck-Aschoff, A., Timmerman, M. E., & Granic, I. (2018). Emotion differentiation and its relation with emotional well-being in adolescents. *Cognition and Emotion*, *32*(3), 651–657. https://doi.org/10.1080/02699931.2017.1338177

Levine, L. J., Lench, H. C., & Safer, M. A. (2009). Functions of remembering and misremembering emotion. *Applied Cognitive Psychology*, *23*(8), 1059–1075. https://doi.org/10.1002/acp.1610

Maciejewski, D. F., van Lier, P. A. C., Branje, S. J. T., Meeus, W. H. J., & Koot, H. M. (2015). A 5-year longitudinal study on mood variability across adolescence using daily diaries. *Child Development*, *86*(6), 1908–1921. https://doi.org/10.1111/cdev.12420

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McArdle, J. J., & Nesselroade, J. R. (1994). Using multivariate data to structure developmental change. In S. H. Cohen & H. W. Reese (Eds.), *The West Virginia University conferences on life-span developmental psychology. Life-span developmental psychology: Methodological contributions* (pp. 223–267). Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.

McLaughlin, K. A., Garrad, M. C., & Somerville, L. H. (2015). What develops during emotional development? A component process approach to identifying sources of psychopathology risk in adolescence. *Dialogues in Clinical Neuroscience*, *17*(4), 403–410. https://doi.org/10.31887/DCNS.2015.17.4/kmclaughlin

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Mill, A., Realo, A., & Allik, J. (2015). Retrospective ratings of emotions: the effects of age, daily tiredness, and personality. *Frontiers in Psychology*, *6*, 2020. https://doi.org/10.3389/fpsyg.2015.02020

Minnaert, A., Prince, A., & Opdenakker, M.-C. (2017). The effect of self-regulated strategy instruction and behavioral consultation on motivation: A longitudinal study on the effect of school-based interventions in secondary education. *Frontiers in Psychology*, *2*(61), 1–15. https://doi.org/10.3389/feduc.2017.00061

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion*, *9*(6), 885–891. https://doi.org/10.1037/a0017823

Mitchell, T. R., & Thompson, L. (1994). A theory of temporal adjustments of the evaluation of events: rosy prospection and rosy retrospection. In C. Stubbart, J. Porac, & J. Meindl (Eds.), *Advances in managerial cognition and organizational information processing* (pp. 85–114). JAI Press.

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view". *Journal of Experimental Social Psychology*, *33*(4), 421–448. https://doi.org/10.1006/jesp.1997.1333

Moè, A., Pazzaglia, F., Tressoldi, P., & Toso, C. (2009). Attitude towards school, motivation, emotions, and academic achievement. In J. E. Larson (Ed.), *Educational psychology: Cognition and learning, individual differences and motivation* (pp. 259–274). Nova Science Publishers.

Moksnes, U. K., Løhre, A., Lillefjell, M., Byrne, D. G., & Haugan, G. (2016). The association between school stress, life satisfaction and depressive symptoms in adolescents: Life satisfaction as a potential mediator. *Social Indicators Research*, *125*(1), 339–357. https://doi.org/10.1007/s11205-014-0842-0

Muthén, L. K., & Muthén, B. (1998–2017). *Mplus user's guide* (8th ed.). https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2020). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology*, *119*(1), 185–203. https://doi.org/10.1037/pspp0000248

Newland, L. A., Giger, J. T., Lawler, M. J., Roh, S., Brockevelt, B. L., & Schweinle, A. (2019). Multilevel analysis of child and adolescents subjective well-being across 14 countries: Child- and country-level predictors. *Child Development*, *90*(2), 395–413. https://doi.org/10.1111/cdev.13134

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2018). The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological Science*, *29*(8), 1346–1357. https://doi.org/10.1177/0956797618773357

Ottenstein, C., & Lischetzke, T. (2020). Recall bias in emotional intensity ratings: Investigating person-level and event-level predictors. *Motivation and Emotion*, *44*, 464–473. https://doi.org/10.1007/s11031-019-09796-4

Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, *21*(4), 331–339. https://doi.org/10.1177/0146167295214003

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–579). Routledge. https://doi.org/10.4324/9780203148211

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International*

handbook of emotions in education (pp. 1–10). Routledge. https://doi.org/10.4324/9780203148211

Riediger, M., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2009). Seeking pleasure and seeking pain: Differences in prohedonic and contra-hedonic motivation from adolescence to old age. *Psychological Science*, *20*(12), 1529–1535. https://doi.org/10.1111/j.1467-9280.2009.02473.x

Riediger, M., Wrzus, C., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2011). Is seeking bad mood cognitively demanding? Contra-hedonic orientation and working-memory capacity in everyday life. *Emotion*, *11*(3), 656–665. https://doi.org/10.1037/a0022756

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Röcke, C., Hoppmann, C. A., & Klumb, P. L. (2011). Correspondence between retrospective and momentary ratings of positive and negative affect in old age: Findings from a one-year measurement burst design. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *66B*(4), 411–415. https://doi.org/10.1093/geronb/gbr024

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS)* [Short scales for measuring positive activation, negative activation and valence in experience sampling studies (PANAVA-KS)] [Research Report]. Department of Psychology, University of Zurich, Switzerland. Retrieved from https://www.psychologie.uzh.ch/dam/jcr:00000000-4a5f-c2e5-0000-000006e4f6b9/PANAVA_05.pdf

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74.

Schimmack, U., & Grob, A. (2000). Dimensional model of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*(4), 325–345. https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I

Schwarz, N. (2012). Why researchers should think "real-time." In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). Guilford Press.

Schweizer, S., Gotlib, I. H., & Blakemore, S.-J. (2020). The role of affective control in emotion regulation during adolescence. *Emotion*, *20*(1), 80–86. https://doi.org/10.1037/emo0000695

Stone, A. A., & Litcher-Kelly, L. (2006). Momentary capture of real-world data. In M. Eid & E. Diener (Eds.), *Multimethod measurement in psychology* (pp. 61–72). American Psychological Association. https://doi.org/10.1037/11383-005

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, *9*, 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510

Tugade, M. M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality*, *72*(6), 1161–1190. https://doi.org/10.1111/j.1467-6494.2004.00294.x

Turner, J. C., & Trucano, M. (2014). Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 643–658). Routledge.

Valois, R. F., Zullig, K. J., Huebner, E. S., & Drane, J. W. (2004). Life satisfaction and suicide among high school adolescents. *Social Indicators Research*, *66*(1/2), 81–105. https://doi.org/10.1023/B:SOCI.0000007499.19430.2f

Venetz, M., & Zurbriggen, C. (2016). Intensity bias oder rosy view? Zur diskrepanz habituell und aktuell berichtetem emotionalen erleben im unterricht [Intensity bias or rosy view? On the discrepancy between self-reported habitual and current experiences in classroom]. *Empirische Pädagogik*, *30*(1), 27–42.

Venetz, M., Tarnutzer, R., Zurbriggen, C., & Sempert, W. (2012). *Emotionales Erleben im Unterricht und schulbezogene Selbstbilder* [Emotional experiences in class and school related self-perceptions]. Edition SZH/CSPS.

Venetz, M., Zurbriggen, C. L. A., Eckhart, M., Schwab, S., & Hessels, M. G. P. (2015). *The Perceptions of Inclusion Questionnaire (PIQ)*. [German Version]. Retrieved from https://piqinfo.ch

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*(5), 820–838. https://doi.org/10.1037/0022-3514.76.5.820

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*(3), 587–600. https://doi.org/10.1111/j.1532-7795.2007.00536.x

Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (pp. 1–74). JAI.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.

Wigfield, A., Byrnes, J. P., & Eccles, J. S. (2006). Development during early and middle adolescence. In P. A. Alexander & H. Winne (Eds.), *Handbook of educational psychology* (pp. 87–113). Erlbaum.

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*(5), 520–524. https://doi.org/10.1111/1467-9280.03455

Wood, J. V., Heimpel, S. A., Manwell, L. A., & Whittington, E. J. (2009). This mood is familiar and I don't deserve to feel better anyway: Mechanisms underlying self-esteem differences in motivation to repair sad moods. *Journal of Personality and Social Psychology*, *96*(2), 363–380. https://doi.org/10.1037/a0012881

Yik, M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, *77*(3), 600–619. https://doi.org/10.1037/0022-3514.77.3.600

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, *11*(4), 705–731. https://doi.org/10.1037/a0023980

Zimmermann, P., & Iwanski, A. (2014). Emotion regulation from early adolescence to emerging adulthood and middle adulthood: Age differences, gender differences, and emotion-specific developmental variations. *International Journal of Behavioral Development*, *38*(2), 182–194. https://doi.org/10.1177/0165025413515405

Zurbriggen, C. L. A., Venetz, M., & Hinni, C. (2018). The quality of experience of students with and without special educational needs in everyday life and when relating to peers. *European Journal of Special Needs Education*, *33*(2), 205–220. https://doi.org/10.1080/08856257.2018.1424777

Zurbriggen, C. L. A., Venetz, M., Schwab, S., & Hessels, M. G. P. (2019). A psychometric analysis of the student version of the Perceptions of Inclusion Questionnaire (PIQ). *European Journal of Psychological Assessment*, *35*(5), 641–649. https://doi.org/10.1027/1015-5759/a000443

Received October 16, 2020
Revision received July 29, 2021
Accepted August 2, 2021 ■