Learning and Instruction 83 (2023) 101626

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

School grades and students' emotions: Longitudinal models of within-person reciprocal effects

Reinhard Pekrun[a,b,c,*], Herbert W. Marsh[b,d], Felix Suessenbach[c], Anne C. Frenzel[c], Thomas Goetz[e]

[a] Department of Psychology, University of Essex, Colchester, United Kingdom
[b] Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia
[c] Department of Psychology, University of Munich, Munich, Germany
[d] Department of Education, Oxford University, Oxford, United Kingdom
[e] Faculty of Psychology, University of Vienna, Vienna, Austria

ARTICLE INFO

Keywords:
Emotion
Grades
Academic achievement
Mathematics
Within-person analysis

ABSTRACT

Based on control-value theory, we expected reciprocal associations between school grades and students' achievement emotions. Existing research has employed between-person designs to examine links between grades and emotions, but has failed to analyze their within-person relations. Reanalyzing data used by Pekrun et al. (2017) for between-person analysis, we investigated within-person relations of students' grades and emotions in mathematics over 5 school years (N = 3,425 German students from the PALMA longitudinal study; 50.0% female). The findings from random-intercept cross-lagged modeling show that grades positively predicted positive emotions within persons over time. These emotions, in turn, positively predicted grades. Grades were negative predictors of negative emotions, and these emotions, in turn, were negative predictors of grades. The within-person effects were largely equivalent to between-person relations of grades and emotions. Implications for theory, future research, and educational practice are discussed.

Feedback about achievement is one of the most important drivers of emotions in achievement settings (Goetz et al., 2018). When you receive information that you have been successful, then you may be happy, proud, and hopeful to again attain success the next time; when you get feedback that you failed, then you may be frustrated, ashamed, and fearful that you might fail again. These effects of feedback have been investigated in studies that used between-person designs. The findings show that success relative to others entails more positive emotions and reduced negative emotions, as compared with the emotions of other students. In addition, the findings also suggest that these emotions, in turn, lead to increased performance relative to the performance of other students.

However, evidence from between-person designs does not inform us about the within-person functional relations that link feedback and emotions. From a theory perspective, feedback sets within-person processes into motion, from perceptions of the feedback to individual interpretations, such as appraisals of control, and emotional responses. Conversely, emotions trigger various cognitive and motivational within-person processes that fuel or hinder subsequent performance. Within-person studies of these reciprocal links are lacking.

The present study aims to fill this crucial gap in the literature. Within-person evidence is needed to more directly test the proposition that feedback on achievement influences students' achievement emotions, and that these emotions reciprocally affect students' achievement. We used data from the longitudinal study of the Project for the Analysis of Learning and Achievement in Mathematics (PALMA) which analyzed student's development in mathematics in secondary school (see, e.g., Murayama et al., 2016; Pekrun et al., 2017, 2019). The dataset included the grades students received in mathematics. Grades are a prime method used to inform students about their achievement in schools around the world. As such, we considered grades as critical indicators of feedback in the present research.

We used the PALMA data to estimate both within-person and between-person relations between grades and emotions, making it possible to examine their equivalence. Findings on relations between grades and emotions in the PALMA longitudinal study were previously reported by Pekrun et al. (2017). However, Pekrun et al.'s analysis used the classic cross-lagged panel model (CLPM), which does not separate

* Corresponding author. Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.
E-mail address: pekrun@lmu.de (R. Pekrun).

https://doi.org/10.1016/j.learninstruc.2022.101626
Received 8 October 2021; Received in revised form 12 April 2022; Accepted 23 April 2022
Available online 21 May 2022
0959-4752/

within-person from between-person effects. In the present analysis, we used the random-intercept cross-lagged panel model (RI-CLPM) to de-confound the two types of relations. This also made it possible to compare the decomposed relations from the present analysis with the undecomposed CLPM findings reported by Pekrun et al. (2017).

1. Theoretical framework: feedback and achievement emotions

We used Pekrun's (2006, 2018, 2021; Pekrun & Perry, 2014) control-value theory (CVT) to derive a theoretical framework on feedback and achievement emotions, with a specific focus on the relation between grades and emotions. We first explain how feedback and achievement emotions are conceptualized in CVT, and then describe theoretical propositions for effects of feedback on emotions, and reciprocal effects of emotions on performance and feedback.

1.1. Feedback

Feedback can be defined as information a system receives about its state. For example, facial feedback is information the central nervous system receives about the action of facial muscles. Using this definition, feedback in education is information students, teachers, or administrators receive about their educationally relevant behaviors and characteristics, such as their learning, teaching, ability, engagement, or social behavior. In educational research using the term, feedback is often more narrowly conceptualized as information students receive about their "performance and understanding" (Hattie & Timperley, 2007, p. 81).

From broader as well more specific definitions, information students receive about their performance is a critical form of feedback. School grades represent information about performance, thus constituting feedback. Grades do not convey much information learners could use to improve their learning. Nevertheless, across education systems, grades are among the most powerful types of feedback given their immediate impact on students' self-evaluations of ability, their long-term consequences for educational and occupational trajectories, and their influence on students' mental health and emotions, as detailed below.

1.2. Achievement emotions

In line with contemporary definitions in the emotion literature (e.g., Scherer & Moors, 2019), CVT views *emotions* as multi-component changes in an organism's psychophysical system that occur in response to important events and actions. These changes can comprise affective, cognitive, physiological, motivational, and expressive-behavioral components. For example, anxiety before an exam can include nervous, uneasy feelings (affective), worries about possible failure (cognitive), physiological arousal (physiological), impulses to avoid taking the exam (motivation), and anxious facial expressions (expressive behavior). CVT defines a*chievement emotions* as emotions that occur in response to events and actions judged according to competence-based standards of quality (Pekrun, 2006). More specifically, achievement emotions are defined as emotions that relate to achievement activities, such as studying, and to their success and failure outcomes. Achievement emotions can be conceptualized as momentary state emotions occurring in a given situation (state achievement emotions), or as trait emotions representing enduring individual dispositions to respond emotionally to achievement situations (trait achievement emotions; see Pekrun, 2006).

The term *affect* is used in emotion research to denote summary constructs that comprise different positive or negative feelings, moods, or emotions. Most often, a binary distinction of positive and negative affect is made, with positive affect denoting a summary construct including positive feelings, and negative affect denoting a summary construct including negative feelings (while disregarding distinctions between different positive and negative feelings; see, e.g., Russell & Barrett, 1999). In the present research, we use the terms positive and negative affect to denote summary constructs including positive and negative achievement emotions, respectively.

1.3. Feedback and achievement emotions: A reciprocal effects model

CVT proposes that feedback about achievement is a prime driver of achievement emotions, due to its impact on appraisals of control and value related to achievement. These emotions, in turn, are thought to impact performance and any subsequent feedback contingent on performance. By implication, feedback and emotions are expected to be linked by reciprocal effects over time.

More specifically, appraisals of control over achievement activities and outcomes, combined with perceptions of their value, are thought to be proximal triggers of achievement emotions. CVT proposes that these emotions are aroused when students feel in control of, or out of control over achievement activities and outcomes that are perceived as important. Positive emotions are typically triggered by a combination of high control and high value, and negative emotions by a combination of low control and high value (for exceptions, see Pekrun, 2006). For example, students will enjoy learning when they feel competent to master the learning material (high control) and are interested in the material (high value), and they will be fearful if they feel incompetent (lack of control) before an exam that is important (high value).

Feedback about achievement is thought to be an especially important factor influencing students' appraisals, thus affecting their achievement emotions (Forsblom et al., 2021; Pekrun, 2018). Positive feedback signaling success is expected to strengthen perceived control and, therefore, to increase positive emotions, such as enjoyment of studying and pride about success. Negative feedback signaling failure undermines perceptions of control, thus exacerbating negative emotions such as anger, anxiety, shame, boredom, and hopelessness.

The feedback implied by grades is deemed to be particularly powerful. In contrast to formative feedback providing information about learning tasks and processes, grades represent summative feedback directed toward the self (Hattie & Timperley, 2007). They can be interpreted as information about ability, thus exerting an enduring influence on perceptions of control. Furthermore, because of their self--relevance, they can also increase perceptions of value. Information about ability and value is especially salient in cumulative grades based on multiple assessments over time, such as end-of-the-year grades derived from multiple single exams. When educational and occupational career opportunities are made contingent on these grades, then the grades signal that achievement is critically important for future trajectories, thus exacerbating related emotions (Pekrun, 2018).

CVT implies that achievement emotions, in turn, impact students' learning, performance, and resulting feedback about performance. Typically, positive emotions strengthen students' intrinsic motivation to learn, task-related attention, use of flexible learning strategies, and self-regulation of learning, thus benefiting achievement. Conversely, negative emotions typically reduce intrinsic motivation, generate task-irrelevant thinking, and undermine flexible strategy use and self-regulation, thus reducing performance. Although positive emotions can occasionally undermine performance (e.g., excessive pride defocusing attention from the task at hand), and negative emotions can sometimes support performance (e.g., anxiety prompting effort to avoid failure), the overall effects of positive emotions are expected to be typically positive, and the overall effects of negative emotions are expected to be typically negative.

In sum, performance feedback and emotions are thought to be linked by reciprocal effects. Feedback interpreted as success prompts positive emotions, which, in turn, generate further success. Feedback interpreted as failure prompts negative emotions, which, in turn, contribute to further failure. Grades are especially likely to be interpreted as success and to drive positive emotions when they represent an improvement over one's average performance, and as failure driving negative emotions when they represent a decline in performance. Similarly, positive and negative emotions may exert especially pronounced effects if they

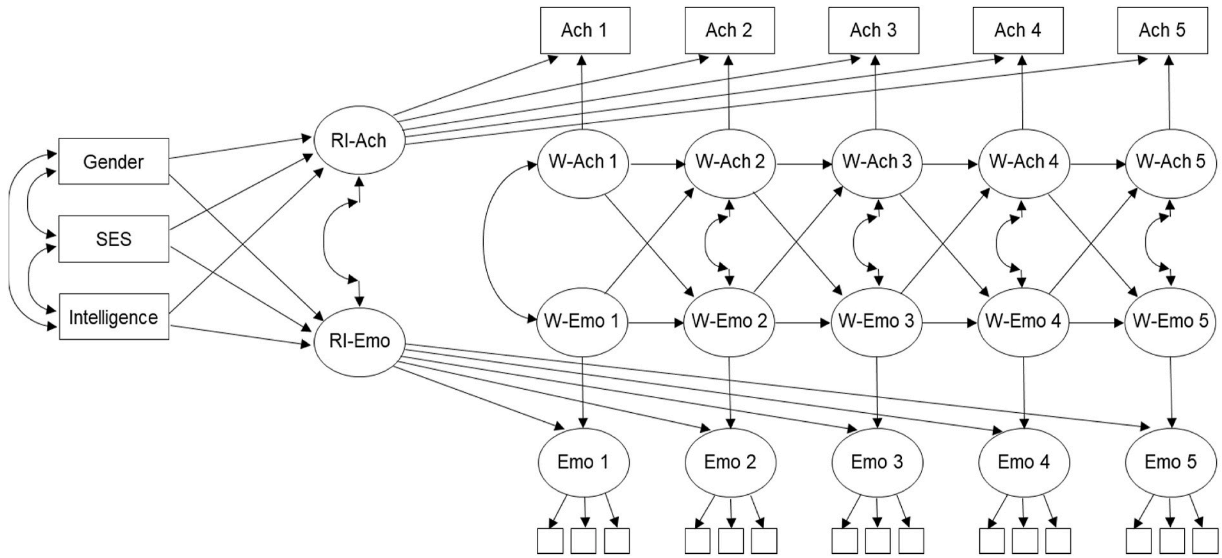

Fig. 1. Random-Intercept Cross-Lagged Panel Model for Grades and Emotion
Note. Ach = achievement feedback (grades). Emo = latent emotion variables. W-Ach = within-person grades factors; W-Emo = within-person emotion factors. RI-Ach and RI-Emo = random intercepts for grades and emotions, respectively. Numbers indicate number of waves.

are stronger than one's average level of positive or negative emotions, respectively. These deviations from the personal baseline are considered in the RI-CLPM used in the present study, as we will detail below (Section 2).

The feedback loops between grades and emotions can extend over different time frames. They can unfold within single achievement situations, such as a single exam; over weeks and months within a school year; or over students' educational career across the years. Whatever the time frame, the effects linking grades and emotions are best conceptualized from a within-person perspective. It is reasonable to assume that the processes linking performance feedback and emotions occur within persons in the first place – success and failure trigger appraisals and emotions within each single student, and these emotions impact each individual student's learning.

Over time, these within-person processes can translate into between-person differences. For example, due to the links between enjoyment and learning, a student who is excited about studying and reaps success will show higher enjoyment and better grades than other students who do not enjoy learning and are less successful. These between-person differences would entail positive correlations between feedback and positive emotions, and negative correlations between feedback and negative emotions, that are equivalent to the respective within-person relations. However, although this proposition seems plausible from a theory perspective, empirical evidence on within-person relations and the possible equivalence of within- and between-person relations for grades and students' emotions is lacking (see Section 3).

2. Methodological approach: within-person and between-person analysis

Educational and psychological research, and research in the social sciences more generally, has typically used between-person empirical designs to test theories of individual functioning. Research on achievement feedback and emotions is no exception from this rule. Nevertheless, the findings are typically interpreted as evidence for within-person psychological mechanisms. Such conclusions may not be warranted. From a statistical perspective, between- and within-person covariances between variables are independent, except if specific conditions hold that are rarely met (ergodicity; Voelkle et al., 2014). Consequently, the findings from between- and within-person analysis can diverge widely (Hamaker et al., 2015; Orth et al., 2021).

As such, it is a critically important task for empirical research to analyze within-person relations more directly and scrutinize their equivalence with between-person findings. Calls for conducting such research have been published in the literature decades ago (for reviews, see Molenaar & Campbell, 2009; Murayama et al., 2017; Pekrun et al., 2002), but have rarely been followed. Fortunately, recent methodological developments facilitate within-person research. Specifically, evolving approaches make it possible to conduct within-person analysis with multi-wave panel data by decomposing within-person from between-person variance (see Usami et al., 2019). Among these approaches, the random-intercept cross-lagged panel model (RI-CLPM) developed by Hamaker et al. (2015) may be especially well suited as it shows fewer problems of model identification and convergence than others (Orth et al., 2021). In the present research, we used this model to disentangle within- and between-person variance in the PALMA dataset, making it possible to estimate within-person and between-person relations of grades and emotions.

3. Prior research on feedback and achievement emotions

Educational research designated to investigate feedback has neglected students' emotional responses, both in theoretical models (see Lipnevich & Panadero, 2021; Panadero & Lipnevich, 2022) and in empirical studies (see Goetz et al., 2018; Lipnevich & Smith, 2009a, 2009b). In Wisniewski et al.'s (2020) meta-analysis of 435 studies on feedback and learning, emotions were not considered. However, several hundred studies investigated relations between school grades and students' emotions (for reviews, see Barroso et al., 2021; Camacho-Morles et al., 2021; Hembree, 1988; von der Embse et al., 2018). Most of these studies aimed to examine links between students' emotions and their performance; as such, grades were considered as an indicator of performance rather than as feedback. Neglect of the feedback function of grades may be a reason why these studies were typically not included in reviews of feedback research. Consistent with our earlier theorizing, the findings document close links between the feedback provided by grades and students' emotions.

The vast majority of research on grades and emotions focused on cross-sectional relations with students' habitual emotional experiences at school. The findings from meta-analyses confirm that grades typically correlated positively with students' habitual positive achievement emotions, and negatively with their habitual negative achievement

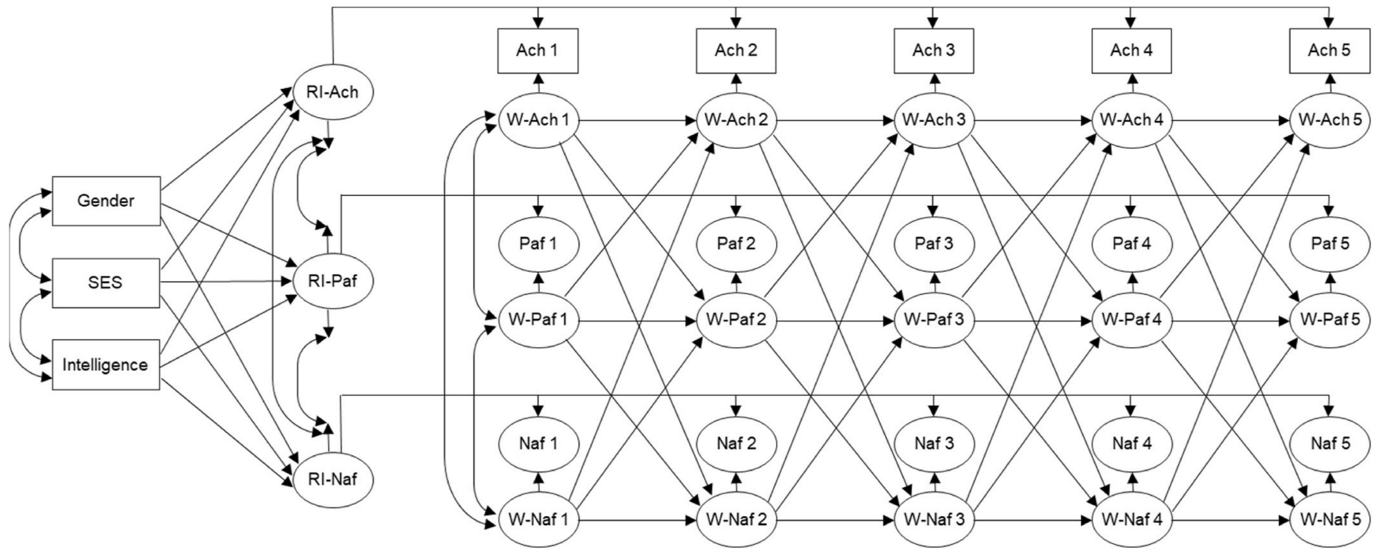

Fig. 2. Random-Intercept Cross-Lagged Panel Model for Grades, Positive Affect, and Negative Affect
Note. Ach = achievement feedback (grades). Paf, Naf = latent positive affect and negative affect variables. W-Ach = within-person grades factors; W-Paf, W-Naf = within-person positive and negative affect factors, respectively. RI-Ach, RI-PAf, RI-Naf = random intercepts for grades, positive affect, and negative affect, respectively. Numbers indicate number of waves. Manifest indicators for latent affect variables and residual covariances are not depicted.

emotions. For studies of grades and learning-related emotions in different school subjects, Camacho-Morles et al. (2021) reported a positive mean true-score correlation for enjoyment (ρ = .30) and negative mean true-score correlations for anger and boredom (ρs = −.35 and −.26, respectively). In the Barroso et al. (2021) meta-analysis, exam and course grades in mathematics showed negative mean correlations with students' math anxiety (*r*s = −.27 and −.20, respectively). Loderer et al. (2020) used meta-analysis to integrate evidence on emotions in technology-enhanced environments. The analysis did not differentiate between grades and test scores, but also reported positive mean correlations of achievement with enjoyment, and negative mean correlations with various negative emotions.

While corroborating that grades and emotions are linked, cross-sectional evidence does not inform us about the directional nature of these links. However, there are a few longitudinal studies that examined the temporal ordering of grades and emotions. All of these studies used between-person designs. Investigations of test anxiety suggested that K-12 students' grades and test anxiety influence each other over time (Meece et al., 1990; Pekrun, 1992; Steinmayr et al., 2016), including negative effects in both directions. In a two-wave path-analytic longitudinal study with 7th to 9th graders, Meece et al. (1990) found that students' math grades predicted their math anxiety. Using the CLPM, Pekrun (1992) showed that students' GPA predicted their test anxiety, and that test anxiety predicted their GPA, over four annual assessments in secondary school (Years 5–8). Steinmayr et al. (2016) reported that 11th grade students' GPA negatively predicted their worry (i.e., the cognitive component of test anxiety) one year later, and that earlier worry predicted later GPA. The effect of GPA on worry was not significant, likely due to the small sample size.

More recent evidence suggests that other emotions can also be reciprocally linked with students' grades. Longitudinal investigations with K-12 students using the CLPM point to reciprocal effects between various emotions and students' grades in mathematics. In the analysis by Pekrun et al. (2017) which used the same PALMA dataset as the present study, students' math grades positively predicted their subsequent positive emotions (enjoyment, pride) and negatively predicted their negative emotions (anger, anxiety, shame, boredom, and hopelessness) over five annual assessments from Years 5–9. In turn, the positive emotions were positive predictors, and the negative emotions were negative predictors of subsequent math grades. In a three-wave study with 5th to 7th graders, Forsblom et al. (2021) found that math grades and students' math emotions (enjoyment, anger, and boredom) were reciprocally related over time. In the four-wave study spanning one school year reported by Putwain et al. (2018), 5th and 6th graders' math grades predicted their enjoyment and boredom in mathematics, and these emotions, in turn, predicted students' math grades.

Similarly, university students' course performance and emotions can be linked by reciprocal effects (Gibbons et al., 2018; Pekrun et al., 2014). In these studies, better grades positively predicted subsequent positive edmotions such as enjoyment, and these emotions, in turn, positively predicted grades. In contrast, poor grades predicted negative emotions such as anger and boredom, and these emotions negatively predicted grades.

However, one fundamental problem with all of these longitudinal studies is the use of analytic models, such as the classic CLPM, that do not disentangle between-person and within-person variance. As argued by Hamaker et al. (2015, p. 102), models such as the CLPM "may lead to erroneous conclusions regarding the presence, predominance, and sign of causal influences," due to their inability to de-confound between- and within-person relations. As demonstrated by these authors with simulated and real datasets, the CLPM may show reciprocal effects that do not exist, fail to detect them when they do exist, and may indicate positive effects when in reality they are negative (or vice versa; for further discussion, see, e.g., Lüdtke & Robitzsch, in press; Marsh et al., in press; Mulder & Hamaker, 2020; Usami et al., 2019).

The existing longitudinal evidence on grades and emotions documents that interindividual distributions of grades are related to interindividual distributions of emotions over time. However, they are not suited to derive conclusions about the within-person causal effects that generate these relations. Within-person analytic designs are needed to answer this question. For school grades and emotions, studies using such designs are lacking. A few studies using experience sampling methodology examined within-person relations between students' appraisals, goals, and emotions (e.g., Goetz et al., 2016; Tanaka & Murayama, 2014), but these studies did not include feedback on achievement or other performance variables.

4. Aims and hypotheses

Given the lack of within-person research on achievement feedback

Table 1
Random-intercept cross-lagged models for grades and emotions: Fit indexes.

	χ^2	*df*	CFI	TLI	RMSEA	SRMR	Factor loadings
Model	*Autoregressive and cross-lagged effects freely estimated*						
Enjoyment	3739.815	1197	.949	.941	.025	.045	.35–.82
Pride	2263.028	763	.958	.950	.024	.038	.56–.78
Anger	2800.137	963	.953	.945	.023	.037	.57–.77
Anxiety	8678.367	3066	.927	.919	.023	.045	.46–.77
Shame	1895.938	963	.974	.970	.017	.030	.53–.76
Boredom	1210.234	570	.981	.976	.018	.029	.61–.76
Hopelessness	1519.892	600	.975	.970	.021	.030	.63–.84
Positive and negative affect	7006.485	729	.946	.933	.050	.073	.41–.96
	Autoregressive and cross-lagged effects invariant across waves						
Enjoyment	3798.788	1209	.948	.940	.025	.047	.35–.82
Pride	2328.788	775	.956	.949	.024	.041	.56–.77
Anger	2855.782	975	.952	.945	.023	.039	.56–.76
Anxiety	8728.257	3078	.926	.918	.023	.046	.46–.77
Shame	1946.671	975	.973	.969	.017	.031	.53–.76
Boredom	1279.217	582	.979	.975	.018	.033	.60–.77
Hopelessness	1580.887	612	.973	.969	.021	.035	.63–.84
Positive and negative affect	7067.504	756	.946	.935	.049	.074	.41–.95

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual. All χ^2 values are significant at $p < .001$.

and student emotions, the first aim of the present study was to use within-person analysis to investigate the relations between feedback and emotions. Feedback was operationalized as students' end-of-the-year grades in mathematics. We examined the links of grades with seven different emotions in math: enjoyment, pride, anger, anxiety, shame, boredom, and hopelessness. Both grades and emotions were assessed annually at the end of each school year from grades 5 to 9, and their relations over time were analyzed using the RI-CLPM. We hypothesized that grades and emotions are reciprocally linked across school years.

Based on our earlier reasoning, we expected the end-of-the-year grades to be a powerful predictor of emotions. In the German education system and similar systems worldwide, end-of-the-year grades represent feedback about students' achievement across the entire year. They represent students' performance across single exams as well as their in-class and homework performance. As such, overall individual performance throughout the year is more precisely reflected in these grades than in any single feedback during the year. Furthermore, important decisions are made contingent on the end-of-the-year grades. Specifically, decisions about repeating grades, skipping grades, and transiting between tracks within the multi-tier German education system are made dependent on these grades. As noted by Roth et al. (2015, p. 118), "school grades are crucial for accessing further scholastic and occupational qualification, and therefore, have an enormous influence on an individual's life."

Due to their salience and importance, the influence of end-of-the-year grades likely extends over a long period of time. Appraisals shaped by these grades influence students' thinking about their abilities and the value of achievement for the year to come (e.g., Marsh et al., 2019). Any subsequent feedback on single exams during the next school year is likely less powerful. Typically, only the summative feedback at the end of the subsequent year can unfold a similar impact. Given the overarching importance of end-of-the-year grades, it is reasonable to assume that they affect not only students' self-beliefs but also their emotions throughout the subsequent school year, until the subsequent year's final grades provide a change (or not).

In terms of the reverse causal direction, we expected students' habitual emotions in mathematics during the school year – as reflected in their self-report towards the end of the year – to influence their learning and performance throughout the year. As such, we expected that these emotions impact the resulting end-of-the-year grades. Based on these assumptions, we expected end-of-the-year grades and emotions to be linked by reciprocal effects from school year to school year.

Our second aim was to use the RI-CLPM to analyze the between-person correlations of grades and emotions, and to examine whether within-person correlations linking the two constructs are equivalent to the between-person correlations. If within-person effects linking two variables translate into equivalent between-person differences over time, then within- and between-person relations can be equivalent, provided that no third variables confound the relations (Goetz et al., 2016). Based on our earlier theoretical reasoning (see Section 1), we tentatively propose within- and between-person relations of grades and emotions to be equivalent at least in terms of occurrence and direction (positive vs. negative). In addition, we also compared the present findings, which decompose within-person and between-person relations, with the undecomposed relations between grades and emotions derived from the CLPM by Pekrun et al. (2017).

Based on CVT and the RI-CLPM, we specified the following preregistered hypotheses for within-person effects and between-person relations linking grades and emotions (https://osf.io/jvt6g/?view_only=4cc3d4a91fa1414eaf9c337d9e9195ee).

Hypothesis 1. The within-person factors for positive emotions and grades are positively correlated, and the within-person factors for negative emotions and grades are negatively correlated.

Hypothesis 2. Similarly, the time-invariant factors (random intercepts) for positive emotions and grades are positively correlated, and the time-invariant factors for negative emotions and grades are negatively correlated.

Hypothesis 3. Grades have positive within-person effects on subsequent positive emotions (enjoyment, pride) and negative within-person effects on subsequent negative emotions (anger, anxiety, shame, boredom, hopelessness).

Hypothesis 4. Positive emotions (enjoyment, pride) have positive within-person effects on subsequent grades, and negative emotions (anger, anxiety, shame, boredom, hopelessness) have negative within-person effects on subsequent grades.

5. Method

5.1. Sample and procedure

The sample consisted of German students who participated in the PALMA longitudinal study (see, e.g., Marsh et al., 2019; Murayama et al., 2016; Pekrun et al., 2017, 2019; for an overview of all PALMA studies, see Pekrun et al., 2007). The PALMA longitudinal study investigated secondary school students' development in mathematics, including their emotions, motivation, self-beliefs, learning strategies,

Table 2
Emotions and grades: Pearson product-moment correlations.

	Enjoyment	Pride	Anger	Anxiety	Shame	Boredom	Hopelessness
	Between-person correlations [a]						
Pride	.796 (.008)	–					
Anger	-.615 (.018)	-.482 (.020)	–				
Anxiety	-.499 (.021)	-.444 (.022)	.811 (.006)	–			
Shame	-.301 (.023)	-.257 (.023)	.714 (.010)	.821 (.006)	–		
Boredom	-.456 (.018)	-.280 (.021)	.619 (.012)	.289 (.016)	.311 (.017)	–	
Hopelessness	-.545 (.019)	-.509 (.020)	.824 (.006)	.929 (.003)	.791 (.007)	.325 (.016)	–
	Within-person correlations [b]						
Pride	.670 (.014)	–					
	.691 (.013)	–					
	.693 (.013)	–					
	.713 (.013)	–					
	.739 (.012)	–					
Anger	-.520 (.024)	-.330 (.024)	–				
	-.539 (.026)	-.425 (.023)	–				
	-.539 (.020)	-.396 (.021)	–				
	-.524 (.021)	-.410 (.027)	–				
	-.491 (.026)	-.404 (.025)	–				
Anxiety	-.388 (.031)	-.298 (.026)	.711 (.013)	–			
	-.388 (.028)	-.357 (.026)	.714 (.011)	–			
	-.364 (.023)	-.323 (.024)	.719 (.011)	–			
	-.354 (.024)	-.305 (.027)	.713 (.014)	–			
	-.333 (.027)	-.314 (.027)	.729 (.011)	–			
Shame	-.241 (.029)	-.184 (.028)	.593 (.019)	.742 (.012)	–		
	-.218 (.025)	-.224 (.027)	.580 (.019)	.741 (.010)	–		
	-.206 (.022)	-.182 (.025)	.540 (.020)	.700 (.013)	–		
	-.171 (.020)	-.176 (.023)	.519 (.020)	.694 (.015)	–		
	-.127 (.027)	-.117 (.028)	.539 (.021)	.705 (.014)	–		
Boredom	-.519 (.020)	-.277 (.022)	.686 (.016)	.434 (.022)	.339 (.023)	–	
	-.515 (.023)	-.335 (.024)	.671 (.016)	.376 (.018)	.291 (.021)	–	
	-.498 (.020)	-.302 (.023)	.642 (.013)	.343 (.018)	.224 (.023)	–	
	-.461 (.019)	-.322 (.022)	.564 (.017)	.295 (.023)	.198 (.027)	–	
	-.412 (.027)	-.303 (.026	.605 (.016)	.347 (.021)	.257 (.023)	–	
Hopelessness	-.365 (.026)	-.323 (.022)	.700 (.015)	.816 (.010)	.705 (.017)	.417 (.024)	–
	-.394 (.023)	-.384 (.023)	.714 (.012)	.839 (.009)	.702 (.014)	.396 (.020)	–
	-.396 (.023)	-.367 (.023)	.714 (.012)	.847 (.006)	.688 (.014)	.362 (.018)	–
	-.399 (.020)	-.371 (.025)	.727 (.013)	.842 (.007)	.669 (.015)	.353 (.019)	–
	-.408 (.024)	-.385 (.026)	.754 (.010)	.859 (.006)	.666 (.017)	.389 (.021)	–
Grades	.125 (.028)	.168 (.028)	-.176 (.023)	-.281 (.020)	-.202 (.020)	-.094 (.023)	-.241 (.022)
	.301 (.016)	.311 (.019)	-.267 (.020)	-.340 (.020)	-.274 (.019)	-.087 (.019)	-.299 (.021)
	.434 (.017)	.395 (.016)	-.328 (.020)	-.355 (.020)	-.215 (.023)	-.173 (.020)	-.362 (.020)
	.514 (.015)	.462 (.016)	-.381 (.015)	-.353 (.021)	-.227 (.017)	-.172 (.018)	-.352 (.021)
	.553 (.015)	.490 (.014)	-.446 (.020)	-.373 (.020)	-.214 (.023)	-.260 (.018)	-.410 (.020)

Note. [a] Between-person correlations are correlations of random intercepts. [b] Within-person correlations are correlations between the within-person centered scores (Hamaker, 2018). 1st, 2nd, 3rd, 4th, and 5th coefficient in each column: Year 5, 6, 7, 8, and 9, respectively. Standard errors (*SE*) in parentheses. $p < .001$ for $r > 2.58$ *SE* (i.e., for all correlations).

and performance in this domain, as well as teacher variables, classroom instruction, and parental support in mathematics. The study included annual assessments from grades 5 to 9. The assessments were administered by the Data Processing Center of the International Association for the Evaluation of Educational Achievement (DPC-IEA) toward the end of each school year. They comprised student, teacher, and parent self--report questionnaires, a mathematics achievement test, and retrieval of grades and demographic information from school documents. Administration of the student instruments was performed in students' classrooms and took three school hours.

Samples were drawn from schools within the state of Bavaria and were representative of the student population of this state in terms of gender, age, types of schools, urban versus rural location, and family background (SES). At the first assessment (grade 5), the sample included 2,070 students from 42 schools (49.6% female, mean age = 11.7 years). In each subsequent year, the study tracked the students who had participated in the previous assessment(s) and recruited additional participants to compensate for attrition (see Pekrun et al., 2007). Sample sizes were 2,059 students in grade 6 (50.0% female, mean age = 12.7 years); 2,397 students at grade 7 (50.1% female, mean age = 13.7 years); 2,410 students at grade 8 (50.5% female, mean age = 14.8 years); and 2,528 students at grade 9 (51.1% female, mean age = 15.6 years). Across all five assessments (i.e., grades 5 to 9), a total of 3,425 students (50.0% female) took part in the study. 38.7% of the total sample participated in all five assessments, and 9.0%, 18.9%, 15.1%, and 18.3% completed four, three, two, or one assessment(s), respectively.

5.2. *Variables and measures*

Grades. Feedback on achievement was measured in terms of the end-of-the-year grades available to students as feedback about their overall math performance during the school year. They represent feedback on achievement relative to the curriculum taught. In Germany, grades

Table 3
Random-intercept cross-lagged panel models: Autoregressive paths, cross-lagged paths, effects of covariates, and correlations of random intercepts.

	Enjoyment model		Pride model		Anger model		Anxiety model		Shame model	
	Enjoyment	Grades	Pride	Grades	Anger	Grades	Anxiety	Grades	Shame	Grades
Autoregressive effects										
T1 → T2	**.398** (.039)	**.306** (.036)	**.366** (.043	**.299** (.038)	**.379** (.035)	**.293** (.039)	**.438** (.030)	**.291** (.040)	**.379** (.031)	**.308** (.040)
T2 → T3	**.402** (.037)	**.308** (.031)	**.395** (.050)	**.301** (.033)	**.407** (.038)	**.293** (.033)	**.444** (.031)	**.289** (.034)	**.407** (.036)	**.307** (.035)
T3 → T4	**.428** (.041)	**.337** (.038)	**.395** (.050	**.330** (.041)	**.436** (.043)	**.321** (.041)	**.489** (.031)	**.315** (.041)	**.389** (.030)	**.335** (.042)
T4 → T5	**.415** (.041)	**.337** (.037)	**.394** (.054	**.330** (.039)	**.420** (.043)	**.321** (.039)	**.475** (.035)	**.318** (.040)	**.407** (.034)	**.336** (.041)
Cross-lagged effects	*Grades → enjoyment*	*Enjoyment → grades*	*Grades → pride*	*Pride → grades*	*Grades → anger*	*Anger → grades*	*Grades→ anxiety*	*Anxiety → grades*	*Grades → shame*	*Shame → grades*
T1 → T2	**.070** (.023)	**.097** (.026)	**.079** (.019)	**.108** (.028)	**-.043** (.019)	**-.112** (.024)	-.031 (.020)	**-.109** (.029)	-.016 (.014)	-.048 (.027)
T2 → T3	**.080** (.026)	**.086** (.023)	**.091** (.021)	**.101** (.028)	**-.047** (.020	**-.110** (.024)	-.032 (.021)	**-.105** (.029)	-.018 (.016)	-.046 (.026)
T3 → T4	**.097** (.032)	**.083** (.024)	**.105** (.025)	**.096** (.028)	**-.055** (.024)	**-.110** (.025)	-.038 (.025)	**-.108** (.029)	-.019 (.018)	-.044 (.025)
T4 → T5	**.103** (.034)	**.075** (.022)	**.105** (.025)	**.092** (.028)	**-.056** (.025)	**-.103** (.024)	-.039 (.026)	**-.103** (.029)	-.020 (.018)	-.044 (.025)
Effects of covariates on random intercepts										
Gender	**.180** (.034)	.023 (.027)	**.218** (.033)	.024 (.026)	**-.067** (.030)	.022 (.027)	**-.212** (.027)	.021 (.026)	-.045 (.028)	.021 (.026)
SES	**-.059** (.028)	**.089** (.021)	**-.064** (.031)	**.088** (.021)	**-.051** (.025)	**.085** (.020)	-.040 (.027)	**.084** (.021)	**-.072** (.027)	**-.084** (.021)
Intelligence	.007 (.039)	**.418** (.036)	-.012 (.035)	**.424** (.036)	**-.186** (.036)	**.426** (.036)	**-.268** (.031)	**.425** (.036)	**-.301** (.035)	**.426** (.036)
	Enjoyment & grades		*Pride & grades*		*Anger & grades*		*Anxiety & grades*		*Shame & grades*	
Correlations of random intercepts	**.468** (.039)		**.424** (.039)		**-.482** (.041)		**-.494** (.052)		**-.431** (.038)	

	Boredom model		Hopelessness model		Positive and negative affect model [a]			
	Boredom	Grades	Hopelessness	Grades	Positive Affect	Negative Affect	Grades	
Autoregressive effects								
T1 → T2	**.439** (.035)	**.309** (.037)	**.331** (.026)	**.292** (.039)	**.566** (.028)	**.465** (.025)	**.275** (.036)	
T2 → T3	**.498** (.031)	**.311** (.032)	**.370** (.031)	**.290** (.034)	**.635** (.031)	**.567** (.024	**.285** (.031)	
T3 → T4	**.595** (.041)	**.338** (.040)	**.384** (.032)	**.318** (.041)	**.695** (.033)	**.689** (.023	**.306** (.037)	
T4 → T5	**.542** (.043)	**.339** (.038)	**.392** (.035)	**.321** (.040)	**.675** (.036)	**.686** (.024)	**.314** (.036)	
Cross-lagged effects	*Grades → boredom*	*Boredom → grades*	*Grades → hopelessness*	*Hopelessness → grades*	*Grades → positive affect*	*Positive affect → grades*	*Grades → negative affect*	*Negative affect → grades*
T1 → T2	-.029 (.018)	**-.097** (.023)	**-.068** (.020)	**-.104** (.024)	**.042** (.012)	**.122** (.024)	**-.026** (.012)	**-.050** (.015)
T2 → T3	-.030 (.019)	**-.105** (.026)	**-.073** (.020)	**-.107** (.025)	**.045** (.012)	**.132** (.026)	**-.024** (.011)	**-.069** (.021)
T3 → T4	-.038 (.023)	**-.110** (.028	**-.082** (.024)	**-.109** (.026)	**.051** (.014)	**.136** (.029)	**-.026** (.012)	**-.082** (.026)
T4 → T5	-.039 (.024)	**-.097** (.026)	**-.086** (.025)	**-.107** (.027)	**.053** (.015)	**.132** (.029)	**-.026** (.012)	**-.083** (.026)
Effects of covariates on random intercepts								
Gender	**.191** (.028)	.021 (.027)	**-.241** (.027)	.023 (.026)	**.193** (.034)	**-.172** (.024)	.020 (.027)	
SES	-.014 (.033)	**.086** (.020)	-.039 (.024)	**.085** (.021)	**-.061** (.026)	**-.046** (.020)	**.093** (.021)	
Intelligence	.041 (.038)	**.428** (.035)	**-.201** (.032)	**.423** (.036)	-.014 (.038)	**-.187** (.024)	**.415** (.037)	
	Boredom & grades		*Hopelessness & grades*		*Positive affect & grades*		*Negative affect & grades*	
Correlations of random intercepts	**-.279** (.065)		**-.516** (.043)		**.415** (.038)		**-.482** (.035)	

Note. Standard errors (*SE*) in parentheses. $p < .05$, .01, and 0.001 for coefficients higher than 1.96, 2.58, and 3.29 *SE*, respectively. **Bold** coefficients: $p < .05$.
[a]Cross-paths between positive and negative affect were not significant (all *ps* > .05).

range from 1 (excellent) to 6 (insufficient). To ease interpretation, we inverted the grade scores in the analysis.

Emotions. We assessed students' emotions in math with the Achievement Emotions Questionnaire-Mathematics (AEQ-M; Pekrun et al., 2011). The instructions ask respondents to indicate how they typically feel when attending class, doing homework, and taking tests and exams in mathematics. The instrument measures mathematics enjoyment (9 items, e.g., "I enjoy my math class"), pride (8 items; e.g., "After a math test, I am proud of myself"), anger (8 items; e.g., "I am annoyed during my math class"), anxiety (15 items; e.g., "I worry if the material is much too difficult for me"), shame (8 items; e.g., "I am ashamed that I cannot answer my math teacher's questions well"), boredom (6 items; e.g., "My math homework bores me to death"), and hopelessness (6 items; e.g., "During the math test, I feel hopeless"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (α range = .83–.92; see Table S1 in Supplementary Materials).

Background variables. Students' gender, family socio-economic status (SES), and intelligence were controlled in the analysis. Gender was coded 1 = *female*, 2 = *male*. SES was assessed by parent report using the EGP classification (Erikson et al., 1979), which consists of six ordered categories of parental occupational status. Higher values represent higher SES. Intelligence was measured at Time 1 (Year 5) using the 25-item nonverbal reasoning subtest of the German adaptation of Thorndike's Cognitive Abilities Test (Kognitiver Fähigkeitstest [KFT 4–12 + R]; Heller & Perleth, 2000).

5.3. Data analysis

We used Hamaker et al.'s (2015) RI-CLPM to test our hypotheses. However, we expanded the classic RI-CLPM (Hamaker et al., 2015) in three important ways (Figs. 1 and 2). First, the original version of the RI-CLPM uses manifest variables to estimate relations, thus not correcting for measurement error. Given that the AEQ-M provides multiple items to measure each emotion, we estimated the emotion variables as latent constructs (see Mulder & Hamaker, 2020, for extending the RI-CLPM by using multiple indicators per construct). Grades were evaluated as manifest variables. Second, following Mulder and Hamaker (2020), we controlled for covariates in the analysis, including the background variables cited above (gender, SES, and intelligence). Third, in a tripartite model including positive and negative affect factors based on multiple emotions, we extended the original two-variable RI-CLPM by including three target variables (i.e., grades, positive affect, and negative affect; Fig. 2). We employed Mplus (Version 8.6; Muthén & Muthén, 1998-2017) to estimate the models.

The RI-CLPM decomposes the target variables into within- and between-person components (i.e., within-person factors and random intercepts). The within-person factors are defined by within-person centering of the variables (Hamaker, 2018). From a multilevel perspective, the within-person components are located at Level 1, and the random intercepts at Level 2. In the present analysis, we used the manifest grade variables and the latent emotion and affect variables as observed variables and decomposed them into within-person factors and random intercepts (Figs. 1 and 2).

Based on this decomposition, the RI-CLPM estimates within-person auto-regressive and cross-lagged relations between the variables and the between-person correlation of the random intercepts. Due to within-person centering, the within-person factors that are used to estimate within-person relations represent the time-specific deviation of a score from the individual's mean score on the respective variable (represented by the random intercepts). The resulting autoregressive and cross-lagged effects represent the relations between these deviations over time. In the present analysis, they answer the question of whether high or low grades (defined relative to the person average) in one year are related to high or low grades and emotion scores (relative to the person average) in a subsequent year. Similarly, they answer the question of whether emotion scores in one year are related to scores on the same emotion as well as grades in a subsequent year, with scores defined relative to the person mean. The random intercepts represent time-invariant global trait factors, thus making it possible to estimate the between-person correlation of the time-invariant components of grades and emotions.

Given multicollinearity between the single emotion constructs (see Section 6), we followed recommendations by Pekrun et al. (2017, 2019) and estimated separate models for the seven different emotions. In addition, following these authors, we estimated one integrative model combining all emotions into two higher-order positive and negative affect factors (tripartite model; Fig. 2). As such, the analysis combines two strategies to deal with multicollinearity, namely, using single variables (separate discrete emotion RI-CLPMs) and combining them by constructing summary variables (integrative affect RI-CLPM).

Measurement models. Prior to estimating the RI-CLPMs, we used confirmatory factor analysis (CFA) to establish measurement models for the emotion and affect variables. To consider the internal structure of the emotion constructs, we used a correlated uniqueness approach by including correlations between residuals for items representing the same setting (attending class, doing homework, and taking tests and exams in mathematics; Pekrun et al., 2011). In addition, correlations between residuals for identical emotion items across measurement occasions were included to control for systematic measurement error.

Subsequently, we constructed an integrative affect measurement model simultaneously representing all emotions. In this model, factor scores derived from the single-emotion CFAs served as indicators. We used this procedure to reduce computational load, but also to support comparison with the Pekrun et al. (2017) findings by directly replicating the strategy used by these authors. Factor scores for the positive and negative emotions served as indicators for positive and negative affect, respectively.

Measurement invariance over time. For each of the emotion and affect constructs, we evaluated models of configural invariance (same patterns of factor loadings), metric invariance (equal factor loadings), scalar invariance (equal loadings and intercepts), and residual invariance (equal loadings, intercepts, and residual variances; Meredith, 1993). For evaluating correlations and path coefficients, metric invariance is the minimum needed (Chen, 2007; Steenkamp & Baumgartner, 1998). To compare model fit, we followed recommendations by Chen (2007) and inspected the loss of fit resulting from imposing invariance constraints (see Supplementary Materials). As recommended, we did not focus on the χ^2 difference test because it is overly sensitive to sample size (Marsh et al., 1988).

RI-CLPMs for discrete emotions and integrative affect factors. Each of the seven discrete emotion models included auto-regressive and cross-lagged effects for within-person factors representing grades and the respective emotion across the five waves (see Fig. 1). The tripartite affect model included auto-regressive and cross-lagged effects linking three variables over time: grades, positive affect, and negative affect (Fig. 2). The background variables were included as covariates in each of the models. These variables represent between-person factors. As such, we included directional paths to the global trait factors (random intercepts) for each of these variables (Mulder & Hamaker, 2020).

Estimator, missing data, and fit indexes. We used robust maximum likelihood (MLR) estimation to deal with non-normality of variables. As students were nested in schools, we corrected for the clustering of the data using the <type = complex> option implemented in Mplus. To handle missing data, we employed full information maximum likelihood (FIML) estimation. FIML has been found to result in trustworthy, unbiased estimates for missing values even in the case of large numbers of missing values (Enders, 2010) and to be an adequate method to manage missing data in studies with large samples (Jeličič et al., 2009).

To evaluate model fit, we applied the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the square root mean residual (SRMR). Traditionally, CFI and TLI values higher than .90 and close to .95, RMSEA values lower than .06, and SRMR values lower than .08 were interpreted as indicating good fit (Hu & Bentler, 1999). However, it should be noted that these recommended cutoff values are often not met with data sets derived from more complex studies, suggesting that they should be used with caution (Marsh et al., 2004).

6. Results

6.1. Measurement models and invariance over time

The configural invariance models showed a good fit to the data, with CFI > .93, TLI > .92, RMSEA < .03, and SRMR < .05 for all seven emotion constructs, and CFI = .957, TLI = .940, RMSEA = .054, and SRMR = .077 for the positive and negative affect model (see overview of findings and Table S2 in Supplementary Materials). Loss of fit for metric, scalar, and residual invariance was negligible, supporting invariance over time and documenting that the constructs met the requirements to be included in the RI-CLPMs.

6.2. Fit of random-intercept cross-lagged panel models

We estimated two sets of models. In the first set, path coefficients were freely estimated. In the second set, we constrained the coefficients to be invariant over time (developmental equilibrium; Marsh et al., 2019). As can be seen from Table 1, constraining the coefficients did not lead to a substantial loss of fit. Because the differences were negligible, we proceeded with the constrained models. The RI-CLPMs for the resulting emotion models fit the data well, with CFI > .92, TLI > .91, RMSEA < .03, and SRMR < .05 for all seven models (Table 1). Similarly, the positive and negative affect model fit the data; CFI = .946, TLI = .935, RMSEA = .049, and SRMR = .074.

6.3. Within-person and between-person correlations

To examine the within- and between-person relations among different emotions, and between grades and emotions, we extracted the within-person factors and the random intercepts for grades and emotions from the RI-CLPMs and calculated their correlations. This analysis involved two steps. We first saved the within-person centered variables and the random intercepts from the RI-CLPMs for the different emotions, and then used Mplus to estimate their correlations. The resulting within-person correlations answer the question of whether within-person deviations from the person average on a given emotion relate to within-person deviations on another emotion (or on grades). The between-person correlations answer the question of whether the trait-like person averages are related.

As shown in Table 2, enjoyment and pride correlated positively, and the correlations between the negative emotions were positive as well. The correlations between positive and negative emotions were negative. This pattern appeared both for the within-person and the between-person correlations. However, the between-person correlations were consistently larger than the within-person correlations.

In addition, enjoyment and pride showed significant positive within-person correlations with grades (range r = .12 to .55; all ps < .001). Anger, anxiety, shame, boredom, and hopelessness showed significant negative within-person correlations with grades (range r = −.17 to −.44; all ps < .001). The between-person correlations of grades and emotions (i.e., the correlations of the random intercepts) showed the same pattern (Table 3). These correlations depict the relations between the time-invariant emotion factors, on the one hand, and the time-invariant grades factor, on the other. Enjoyment and pride showed strong positive correlations with grades (rs = .46 and .42, respectively; ps < .001), and all negative emotions showed negative correlations with grades (range r = −.27 to −.51; all ps < .001).

6.4. Autoregressive and cross-lagged effects

All of the emotion and affect constructs showed substantial autoregressive within-person stability over time (Table 3). The range of standardized autoregressive coefficients for the seven emotion variables was ß = .33 to .59, and for the positive and negative affect constructs .46 to .69 (all ps < .001). Similarly, there was moderate within-person stability for the math grades, ß range = .29 to .33 (ps < .001). These coefficients indicate that deviations in grades and emotions from the individual person average were positively related to deviations in the same variables at the next wave.

In addition, there were significant cross-lagged effects for most of the emotions. These effects show that the within-person fluctuations of grades and emotions were associated over time. The effects of grades on emotions indicate that deviations of grades from the individual average grade were related to subsequent deviations of emotions from the individual average emotion scores. The effects of emotions on grades indicate that deviations in the emotions were associated with subsequent deviations in the grades.

More specifically, grades had positive predictive effects on subsequent enjoyment and pride, and enjoyment and pride had positive effects on subsequent grades. Grades had significant negative effects on subsequent anger and hopelessness; anger, anxiety, boredom, and hopelessness had significant negative effects on subsequent grades. The effects of grades on anxiety, shame, and boredom, and the effects of shame on grades were not significant. However, the direction of these effects was negative as well. Finally, grades and positive affect were linked by significant positive reciprocal effects over time, and grades and negative affect were linked by significant negative reciprocal effects.

6.5. Effects of covariates

Gender had significant positive effects on enjoyment, pride, and positive affect, and significant negative effects on anger, anxiety, boredom, hopelessness, and negative affect (Table 3; effects of covariates on random intercepts). These coefficients indicate that male students reported more positive emotions and less negative emotions in math than female students. Gender did not show any significant effects on students' grades in math. Intelligence and family SES related positively to students' grades. In addition, intelligence was a negative predictor of students' negative emotions.

6.6. Comparison with CLPM findings

The current RI-CLPMs and the CLPMs reported by Pekrun et al. (2017) rendered equivalent patterns of cross-lagged effects and grade-emotion correlations. In both modeling approaches, grades related positively to subsequent positive emotions and negatively to subsequent negative emotions (although not all of these paths were significant in the RI-CLPMs). Conversely, positive emotions related positively, and negative emotions related negatively to subsequent grades. Furthermore, the within-person and between-person correlations of grades and emotions derived from the RI-CLPM were equivalent to the between-person correlations for the single waves reported by Pekrun et al. (2017). However, the RI-CLPM between-person correlations were stronger than the single-wave between-person correlations. The average (median) of the RI-CLPM correlations between random intercepts was $|r|$ = 0.46 (Table 3). The average (median) of the single-wave between-person correlations was $|0.34|$ (Pekrun et al., 2017, Table 1).

7. Discussion

The present study examined within-person relations between teacher feedback on students' achievement, as represented by school grades, and students' emotions over five school years. Based on control-value theory, we aimed to test the hypothesis that grades influence the development of students' emotions, and that these emotions, in turn, impact the grades students receive. To the extent that grades mirror students' performance, the analysis also reveals how performance and emotions are related over time, from a within-person perspective. Such within-person evidence is needed to understand the effects of feedback, to adequately plan interventions, and to improve educational practices.

A second aim was to compare the findings with between-person relations linking grades and emotions. The relative equivalence of within- and between-person relations is of prime relevance from both substantive and methodological perspectives. If these relations are equivalent, then there is support for the position that variables function in similar ways at the within- and between-person levels, and common theoretical principles can be used to explain relations at the two levels. If this is not the case, then different theories are needed to explain within- and between-person relations of the variables.

The present analysis makes it possible to inspect the equivalence of within- and between-person findings from three perspectives. First, we can directly compare the estimates for within- and between-person relations in the present RI-CLPMs. Second, because we used the same dataset as Pekrun et al. (2017), we can compare the current results with the undecomposed CLPM findings reported by these authors. Third, we can compare them with between-person findings reported for other datasets in the extant literature. In the following, we first discuss the findings for relations among the emotions and for their links with grades, and then consider limitations and future directions.

7.1. Relations among emotions

The pattern of relations among the seven emotions was the same across the within- and between-person levels. Without exception, as derived from the present RI-CLPMs, there were positive correlations within the group of positive emotions, positive correlations within the group of negative emotions, and negative correlations between the two groups of emotions. These correlations are also equivalent to the between-person correlations reported for each single wave by Pekrun et al. (2017). As such, there seems to be equivalence in the relations between students' achievement emotions at the within- and between-person levels, at least in terms of the occurrence, significance, and direction of associations.

Between-person relations among emotions can result from stable factors that vary between persons, such as genetic dispositions (Miu et al., 2019). However, at least for achievement emotions the present findings suggest that this is not the whole story. Given that the current relations among emotions were also observed at the within-person level, they cannot be explained by between-person factors alone. Rather, factors that vary within persons are needed to explain the within-person covariance of emotions. Factors that covary with emotions in similar ways at the within- and between-person levels could explain equivalence, such as personal beliefs, individual perceptions, or grades as examined in the present study.

For example, between-person correlations show that students' enjoyment and pride are positively associated with antecedent competence beliefs (e.g., Pekrun et al., 2019). These positive associations help explain why the between-person correlation of enjoyment and pride is typically positive. Similarly, it is reasonable to expect that any within-person fluctuation of enjoyment and pride relates positively to the within-person fluctuation of competence perceptions – learning is more enjoyable, and makes us feel proud, when feeling competent in a given situation. The positive within-person relations to competence perceptions help explain why the within-person correlation of enjoyment and pride is positive. In conclusion, due to the equivalence of between- and within-person links to antecedent competence perceptions, the relations between the two emotions may also be equivalent across the two levels.

7.2. Grades and emotions

Correlations. Supporting Hypotheses 1 and 2, grades correlated positively with positive emotions and positive affect, and negatively with negative emotions and negative affect. Again, these correlations were equivalent across the two levels in terms of significance and direction, and they were of substantial magnitude. Supporting our hypotheses, the between-person correlations with grades were $|r| > .42$ for all emotions except boredom (Table 3). At the within-person level, the correlations ranged from $r = .12$ to .55 (Table 2).

The between-person correlations can be compared with average between-person correlations reported in the literature. As described earlier (Section 4), mean true-score correlations of grades with enjoyment, anger, and boredom were $\rho s = .30, -.35$, and $-.26$, respectively, in the analysis by Camacho-Morles et al. (2021), and mean correlations of math grades and math anxiety ranged from $r = -.20$ to $-.27$ in the analysis by Barroso et al. (2021). In addition, the present findings can be compared with the between-person correlations based on the PAMA dataset reported by Pekrun et al. (2017, Table 1; average correlation $|r| = .34$).

In conclusion, on average, the present decomposed between-person correlations were higher than average undecomposed correlations as reported in the literature or derived from the same dataset. Undecomposed correlations typically suggest that emotions explain around 10% of the variance in grades, and vice versa. The present correlations suggest that a more realistic estimate may be around 20% explained variance, provided that variances are decomposed into between- and within-person components. In fact, the magnitude of the present decomposed between-person correlations is similar to typical correlations between students' cognitive abilities and their grades (rs around 0.45 for intelligence and school grades; see the meta-analyses in Kriegbaum et al., 2018; Roth et al., 2015).

Stability over time. Students' grades as well as all of the emotion and affect variables showed significant positive within-person autoregressive effects over time. The range of effects across models was $\beta = .29$ to .33 for grades, .29 to .59 for the emotions, and .46 to .69 for positive and negative affect. These effects were smaller than the undecomposed autoregressive effects reported by Pekrun et al. (2017, Table 3), but still substantial.

The effects indicate that positive deviations from the individual person average in one school year tend to be followed by a positive deviation in the next school year, and negative deviations by a negative deviation. This is not a trivial finding. Alternatively, the within-person deviations might have been independent across years, or even negatively related in terms of scores bouncing back and forth over time. Instead, the positive effects suggest that there are positive carry-over effects (i.e., inertia) from year to year, implying that both grades and emotions tend to persist over time before returning to the person average.

Reciprocal effects. The findings for cross-lagged effects linking grades and emotions largely supported our hypotheses. Over and above autoregressive and reciprocal effects, good grades positively predicted students' subsequent positive emotions and negatively predicted their subsequent negative emotions, in line with Hypothesis 3. Positive emotions, in turn, were positive predictors of students' subsequent grades, and negative emotions were negative predictors, supporting Hypothesis 4. Most of these effects were significant, and all of them were consistent in terms of their direction. As such, the findings corroborate the importance of grades as a driver of students' emotions, and of emotions as drivers of students' learning and achievement. The two sets of predictive effects combined support reciprocal effects models of feedback and achievement, on the one hand, and emotions, on the other (e.g., Pekrun, 1992; Pekrun et al., 2017).

Importantly, similar to correlations and autoregressive effects, the within-person path coefficients were largely equivalent to the between-person effects reported in the literature, and to the undecomposed path coefficients reported by Pekrun et al. (2017) for the PALMA dataset. Similar to the present within-person findings, the coefficients reported by Pekrun et al. (2017) were positive for effects linking grades with positive emotions, and negative for effects linking them with negative emotions. As such, there is convergence of the current within-person reciprocal effects and Pekrun et al.'s (2017) undecomposed effects linking grades and emotions, at least in terms of the direction of effects.

The existing longitudinal studies of achievement and emotions cited earlier, including Pekrun et al.'s (2017) analysis, used cross-lagged panel modeling (CLPM) not decomposing within- and between-person variance. As such, the present and previous findings combined suggest directional equivalence of achievement-emotion associations across three analytically different types of associations: (1) relations of the time-varying within-person factors estimated in the RI-CLPM; (2) between-person relations of the time-invariant random intercepts from the RI-CLPM; and (3) the undecomposed relations estimated in the CLPM.

Given that findings based on within-person and between-person analysis often diverge (Hamaker et al., 2015; Orth et al., 2021), the degree of equivalence in the present data is remarkable. However, from a theory perspective, it is plausible that within- and between-person relations show equivalence more often than might be expected on statistical grounds alone. Specifically, as argued earlier, it seems sensible that within-person processes can translate into equivalent between-person differences over time. For example, if positive teacher feedback repeatedly triggers positive achievement emotions in a student, then this cumulative feedback entails that the student develops stronger trait positive achievement emotions over the years than other students. Conversely, if positive emotions continuously support a student's learning, then the student can develop stronger competencies and habitually outperform others. As long as the resulting between-person differences are not disturbed by other factors, they should be equivalent to the underlying within-person dynamics.

Similarly, between-person differences can trigger equivalent within-person processes. For example, trait emotions that correlate positively with achievement can trigger positive state emotions that strengthen a student's current performance, thus also leading to equivalence of within- and between-person relations. In sum, within-person processes can result in equivalent between-person differences, and these differences, in turn, can prompt equivalent within-person processes (see also Wrzus et al., 2021).

7.3. Limitations and directions for future research

The present study used longitudinal data and latent modeling controlling for covariates, and it yielded robust within-person and between-person evidence. Nevertheless, there are several limitations to the study that should be kept in mind and can be used to plan directions for future research.

First, an especially important issue for examining within-person effects is the length of the time intervals between assessments. In the current analysis, the assessments were one year apart. As such, the findings represent the long-term dynamics of grades and emotions over several years. There may be substantial dynamics within shorter intervals as well – within single lessons, days, weeks, and months. In future studies, assessing short-term dynamics may further help to elucidate how within-person processes translate into the associations linking grades and emotions. The size of the correlations between grades and emotions (Tables 2 and 3) suggests that there is more variance to be explained in subsequent studies, over and above the year-to-year cross-lagged effects that we observed in the present analysis.

Second, a related issue is type of feedback (Lipnevich & Smith, 2018). End-of-the-year grades provide students with summative feedback about their overall performance throughout the year. Furthermore, as they are typically defined, they inform students about their performance relative to other students. As such, they are of limited value to assess individual improvement, and they do not inform students about how to strengthen their future learning. It seems likely that other types of feedback influence students' emotions as well, and possibly in better ways. For future studies, it may be especially important to examine effects of informational feedback about learning, as well as effects of daily feedback provided in teacher-student interaction (e.g., teacher comments involving praise and blame during lessons).

Third, to interpret the present findings, it is important to consider how the RI-CLPM defines relations at the within-person level. As noted, these relations answer the question of how within-person deviations from the person average on different variables are related. More specifically, they represent covariation of the between-person distributions of these within-person deviations. They do not represent idiographic associations between variables within single persons. Investigating feedback-emotion relations for single persons, and the variation of these relations across different persons, is an important task for future research (e.g., using dynamic structural equation modeling; Asparouhov et al., 2018). Given the limited number of school years, it would be difficult to do this with end-of-the-year grades. However, it should be possible to design such an analysis using multiple grades for single exams (see Schultzberg & Muthén, 2018, for numbers of waves needed).

Furthermore, we analyzed molar relations between students' grades and their emotions but did not examine mediating mechanisms. CVT suggests that an important mechanism explaining effects of feedback on achievement emotions is students' perceptions of control and value. Mediation of achievement effects on emotions by appraisals has been explored in between-person research (Forsblom et al., 2021). Future studies should examine the within-person mechanisms that explain the links between feedback and emotions, as well as moderators that could modify these links. An examination of mechanisms and moderators may also help to uncover possible differences in the effects of grades on different emotions, and different emotions on grades.

For example, given that grades represent achievement outcomes, they may have more immediate effects on outcome emotions than on activity emotions. In the present data, relations with outcome emotions (pride, anxiety, shame, hopelessness) were of similar magnitude as relations with activity emotions (enjoyment, anger, boredom). An analysis of mediators and moderators may help explain under what conditions, and in which students, grades exert differential effects on distinct emotions. Similarly, CVT suggests that different emotions influence learning differently (e.g., variable effects of activating negative emotions like anxiety, but uniformly adverse effects of deactivating negative emotions like hopelessness; Pekrun, 2006). Molar relations as analyzed in the present research may mask these nuanced influences. More fine-grained studies that include mediating mechanisms and use designs with higher temporal granularity are needed to test these possible differences.

Finally, the present findings relate to grades and emotions in a specific domain, age group, and country. The principles of relative universality that are part of CVT (Pekrun, 2009, 2018) suggest that the current relations between grades and emotions should also be observed in other academic domains, student groups, and cultural contexts. However, empirically the generalizability of the present findings remains open to question. It may be especially important to conduct within-person research on achievement emotions with younger students; the early school years may be critical in laying the foundations for students' emotions and their links with achievement.

7.4. Implications for educational practice

For educational practice, the present results further confirm and expand the existing findings from between-person research. Specifically, two messages follow from our results. First, within- and between-person findings converge in showing that summative feedback using normative standards as implied by grades can help boost students' positive emotions, but can also exacerbate their negative emotions. As such, use of grades can undermine many students' psychological health. Alternative ways of providing feedback that better support mental health should be explored and practiced (see also Linnenbrink et al., 2016).

Informational feedback that uses mastery standards may be better suited to support all students in developing favorable competence beliefs, thus promoting adaptive emotions and reducing maladaptive emotions. Furthermore, in many education systems educational opportunities are currently made contingent on grades, thus exacerbating their emotional load, as argued earlier. Reducing these contingencies may help prevent excessive perceptions of value, thus reducing emotions that jeopardize mental health, such as hopelessness and excessive test anxiety. Restructuring education systems in this way may incur costs but may also generate benefits for both mental health and student performance (see, e.g., Parker et al., 2018).

Second, the findings suggest that students' emotions are not only important for their wellbeing but also influence their learning and performance (as reflected in grades). As such, teachers, parents, administrators, and policymakers are well advised to consider student emotions in planning and practicing education. Teachers can support students' emotional development by providing high-quality instruction that is clearly structured, personalized in terms of adapting task demands and expectations to students' competencies, and suited to promote students' sense of autonomy and relatedness (Linnenbrink-Garcia et al., 2016; Pekrun, 2014). In addition, teachers can support students by developing and displaying their own positive emotions (Frenzel et al., 2018), by creating mastery goal structures in the classroom, and by scaffolding students' regulation of emotions.

Similarly, parents can create a mastery-oriented instructional climate at home and convey educational aspirations that do not over-challenge students (Murayama et al., 2016). Finally, administrators and policymakers can contribute by integrating emotion-oriented courses in teacher education programs, and by redesigning school curricula to include emotional learning that supports students' development of emotional competencies.

Author statement

Reinhard Pekrun: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Writing - Original Draft, Writing – Review & Editing, Supervision.

Herbert W. Marsh: Fomal Analysis, Writing – Review & Editing.

Felix Suessenbach: Fomal Analysis, Writing – Review & Editing.

Anne C. Frenzel: Writing – Review & Editing.

Thomas Goetz: Writing – Review & Editing.

Declaration of competing interest

We have no conflicts of interest to disclose. This research was supported by four grants from the German Research Foundation (Deutsche Forschungsgemeinschaft) awarded to Reinhard Pekrun (PE 320/11-1, PE 320/11-2, PE 320/11-2, PE 320/11-2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2022.101626.

References

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134–168. https://doi.org/10.1037/bul000030

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483–493. https://doi.org/10.1016/j.jad.2017.11.048

Enders, C. K. (2010). *Applied missing data analysis*. Guilford.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European Societies: England, France, and Sweden. *British Journal of Sociology, 30*, 341–415. https://doi.org/10.2307/589632

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2021). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000671. Advance online publication.

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology, 110*(5), 628–639. https://doi.org/10.1037/edu0000228

Gibbons, R. E., Xu, X., Villafañe, S. M., & Raker, J. R. (2018). Testing a reciprocal causation model between anxiety, enjoyment and academic performance in postsecondary organic chemistry. *Educational Psychology, 38*, 838–856. https://doi.org/10.1080/01443410.2018.1447649

Goetz, T., Lipnevich, A. A., Krannich, M., & Gogol, K. (2018). Performance feedback and emotions. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 554–574). Cambridge University Press.

Goetz, T., Sticca, F., Pekrun, R., Murayama, K., & Elliot, A. J. (2016). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction, 41*, 115–125. https://doi.org/10.1016/j.learninstruc.2015. 10.007

Hamaker, E. L. (2018, March). In *How to run the RI-CLPM with Mplus*. https://www.statmodel.com/download/RI-CLPM Hamaker input.pdf.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. https://doi.org/10.1037/a0038889

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4–12 + R) [Cognitive Abilities Test for Grades 4 to 12*. Hogrefe. revision (KFT 4–12 + R)].

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology, 45*(4), 1195–1199.

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive motivation and emotion in education: Research and principles for instructional design. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 228–236. https://doi.org/10.1177/2372732216644450

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions and conclusions. *Frontiers in Education, 6*, 720195. https://doi.org/10.3389/feduc.2021.720195

Lipnevich, A. A., & Smith, J. K. (2009a). "I really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability, 21*(4), 347–367. https://doi.org/10.1007/s11092-009-9082-2

Lipnevich, A. A., & Smith, J. K. (2009b). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15*(4), 319–333. https://doi.org/10.1037/a0017841

Lipnevich, A. A., & Smith, J. K. (Eds.). (2018). *The Cambridge handbook of instructional feedback*. Cambridge University Press.

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction, 70*, Article 101162. https://doi.org/10.1016/j.learninstruc.2019.101272

Lüdtke, O., & Robitzsch, A. (in press). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling*.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*(3), 391–410. https://doi.org/10.1037/0033-2909.103.3.391

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fix indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., Pekrun, R., & Lüdtke, O. (in press). Directional ordering of self-concept, school grades, and standardized tests over five years: New tripartite models juxtaposing within- and between-person perspectives. *Educational Psychology Review*.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60–70. https://doi.org/10.1037/0022-0663.82.1.60

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/bf02294825

Miu, A. C., Homberg, J. R., & Lesch, K.-P. (2019). In *Genes, brain, and emotions: Interdisciplinary and translational perspectives*. Oxford University Press.

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Mulder, J. D., & Hamaker, E. L. (2020). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 1–11. https://doi.org/10.1080/10705511.2020.1784738

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus user's guide. 8th ed..

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. In D. W. Putwain, & K. Smart (Eds.), *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education – Current Trends: The Role of Competence Beliefs in Teaching and Learning* (pp. 71–87). Wiley.

Murayama, K., Pekrun, R., Suzuki, M., Marsh, H. W., & Lichtenfeld, S. (2016). Don't aim too high for your kids: Parental over-aspiration undermines students' learning in mathematics. *Journal of Personality and Social Psychology, 111*(5), 166–179. https://doi.org/10.1037/pspp0000079

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology, 120*(4), 1013–1034. https://doi.org/10.1037/pspp0000358

Panadero, E., & Lipnevich, A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 10046. https://doi.org/10.1016/j.edurev.2021.100416

Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal, 55*(4), 836–858. https://doi.org/10.3102/0002831218760213

Pekrun, R. (1992). The expectancy-value theory of anxiety: Overview and implications. In D. G. Forgays, T. Sosnowski, & K. Wrzesniewski (Eds.), *Anxiety: Recent developments in self-appraisal, psychophysiological and health research* (pp. 23–41). Hemisphere Publishing.

Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2009). Global and local perspectives on human affect: Implications of the control-value theory of achievement emotions. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 97–115). Hogrefe.

Pekrun, R. (2014). *Emotions and learning* (Educational Practices Series, Vol. 24). Geneva, Switzerland: International Academy of Education (IAE) and International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from http://www.iaoed.org/downloads/edu-practices_24_eng.pdf.

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. In G. A. D. Liem, & D. M. McInerney (Eds.), *Big theories revisited 2: A volume of research on sociocultural influences on motivation and learning* (pp. 162–190). Information Age Publishing.

Pekrun, R. (2021). Self-appraisals and emotions: A generalized control-value approach. In T. Dicke, F. Guay, H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self – a multidisciplinary concept* (pp. 1–30). Information Age Publishing.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Hofmann, H. (2002). Mythen in der Schulleistungs- und Jugendforschung: Möglichkeiten und Grenzen populationsorientierter Untersuchungen [Myths in research on adolescence and achievement: Strengths and limitations of population-based designs]. In H. Merkens, & J. Zinnecker (Eds.), *Jahrbuch Jugendforschung 2/2002* (pp. 127–144). Leske & Budrich.

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology, 106*(3), 696–710. https://doi.org/10.1037/a0036006

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology, 117*(1), 166–185. https://doi.org/10.1037/pspp0000230

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Taylor & Francis.

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17–37). Waxmann.

Putwain, D. W., Becker, S., Symes, W., & Pekrun, R. (2018). Reciprocal relations between students' academic enjoyment, boredom, and achievement over time. *Learning and Instruction, 54*, 73–81. https://doi.org/10.1016/j.learninstruc.2017.08.004

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology, 7*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–90. https://doi.org/10.1086/209528

Steinmayr, R., Crede1, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology, 6*, Article 1994r. https://doi.org/10.3389/fpsyg.2015.01994

Tanaka, A., & Murayama, K. (2014). Within-person analyses of situational interest and boredom: Interactions between task-specific perceptions and achievement goals. *Journal of Educational Psychology, 106*(4), 1122–1134. https://doi.org/10.1037/a0036659

Usami, S., Todo, N., & Murayama, K. (2019). Modeling reciprocal effects in medical research: Critical discussion on the current practices and potential alternative models. *PLoS One, 14*(9), Article e0209133. https://doi.org/10.1371/journal.pone.0209133

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research, 49*(3), 193–213. https://doi.org/10.1080/00273171.2014.889593

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Wrzus, C., Luong, G., Wagner, G. G., & Riediger, M. (2021). Longitudinal coupling of momentary stress reactivity and trait neuroticism: Specificity of states, traits, and age period. *Journal of Personality and Social Psychology, 121*(3), 691–706. https://doi.org/10.1037/pspp0000308